DOCUMENT RESUME

ED 218 124 SE 038 234

AUTHOR Schoenfeld, Alan H.
TITLE Expert and Novice Mathematical Problem Solving. Final Project Report and Appendices B-H.
INSTITUTION Hamilton Coll., Clinton, N.Y.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 82
GRANT SED-79-19049
NOTE 258p.; Appendix A removed because of copyright restrictions. It appears in "Journal for Research in Mathematics Education," Volume 13, number 1, p 31-49, Jan 82.

EDRS PRICE MF01/PC11 Plus Postage.
DESCRIPTORS *Cognitive Processes; *College Mathematics; Educational Research; Higher Education; Learning Theories; *Mathematics Education; *Mathematics Instruction; *Problem Solving; Teaching Methods
IDENTIFIERS *Mathematics Education Research

ABSTRACT

The project consisted of three related series of studies, designed to (a) determine some of the underlying competencies that contribute to experts' successful problem solving performance in college level mathematics, (b) to determine what productive behaviors students lack, or what counterproductive behaviors they have, that keep them from being effective problem solvers, and (c) to determine if a course in mathematical problem solving that explicitly teaches problem solving strategies could significantly improve students' problem solving performance. A variety of new methodologies and measurement techniques were developed for examining cognitive processes in broad domains such as "general mathematical problem solving." Experts were shown to perceive the "deep structure" of problems where novices were misled by the "surface structure." Experts have much better "executive" or metacognitive strategies than novices, which prevent them from squandering their problem solving resources in the way that students do. Moreover, it was shown that students' "belief systems" about mathematics and the way it is done often make it difficult for them to learn mathematics or to use it effectively. The problem solving course provided clear evidence that, with direct instruction, students' problem solving performance could be substantially improved. These results have obvious implications both for research and teaching in mathematical thinking.

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218124

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Expert and Novice Mathematical Problem Solving

National Science Foundation Research in Science Education

Grant SED 7[illegible] - 19049

FINAL REPORT, MAY 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Alan H. Schoenfeld
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Alan H. Schoenfeld, Principal Investigator*
Hamilton College
Clinton, NY 13323

*Current address:
The University of Rochester
Rochester, NY 14627

SE 038 234

Contents

Final Project Report form 98A.....i

Cover.....ii

Contents.....iii

Overview.....1

Major Results.....3

Publication Citations.....9

Scientific Collaborators.....10

Appendices

Appendix A: Measures of Problem Solving Performance and of Problem Solving Instruction

Appendix B: Problem Perception and Knowledge Structure in Expert and Novice Mathematical Problem Solvers

Appendix C: Toward a Testable Theory of Problem Solving

Appendix D: Recent Advances in Mathematics Education: Ideas and Implications

Appendix E: Episodes and Executive Decisions in Mathematical Problem Solving

Appendix F: On the Analysis of Two-Person Problem Solving Protocols

Appendix G: Beyond the Purely Cognitive: Metacognition and Social Cognition as Driving Forces in Intellectual Performance

Appendix H: Some Thoughts on Problem Solving Research and Mathematics Education

Final Project Report
NSF RISE SED 79-19049
Expert and Novice
Mathematical Problem Solving

Overview

This project consisted of three interrelated series of studies examining the nature of expert and novice problem solving processes in college level mathematics. The studies fell into these three categories:

a. A detailed characterization of "expert" mathematical problem solving, and an elaboration of the mechanisms by which experienced problem solvers (college faculty) succeed in solving difficult and sometimes unfamiliar problems.

b. A characterization of "novice" (college freshman) mathematical problem solving behavior, and a comparison of the procedures used by the students with those employed by the faculty. What productive behaviors do the students lack, and what counterproductive behaviors do they exhibit, that keep them from being effective problem solvers?

c. An examination of the effects of a month-long intensive course in mathematical problem solving, based on the research described in (a) and (b) above, on students' problem solving performance. Can explicit instruction in mathematical problem solving processes result in improved performance, even on problems unrelated to those used in the instruction?

Common to all three categories, and an essential notion underlying all of the research, is a focus on the process of mathematical problem solving. This involves the detailed examination of the kinds of reasoning used by individuals as they are involved in the solution of mathematical problems, not simply the measurement of "before" and "after" results. At the time this research was undertaken there was little available in terms of extant

theory, or of methodologies for rigorously examining such complex cognitive processes in broad domains such as "general mathematical problem solving;" there were virtually no measures available for capturing and evaluating such processes. Thus a major component of the research was the development of research methodologies and measurement tools that would serve to characterize such processes, and that would serve as a means of elaborating a more complete theory of intellectual performance in broad, complex domains.

I am pleased to report that all aspects of the project have been completed on time and well under budget; all have more than met the goals outlined in the proposal. The major results are described in the following section, with specific references to documents produced by the project.

Major Results

As I indicated in the annual report, December 1980, the first of the studies scheduled for completion, and the first completed, was study 3. The evaluation of students' problem solving performance before and after the intensive problem solving course called for the development of a series of new measures of problem solving, and it was originally planned that two years would be devoted to the development and refinement of the tests. The development and pilot testing of the measures went very smoothly, and a second year was not needed. The results of study 3 were written up and published in "Measures of problem solving performance and of problem solving instruction," (Appendix A) which was published in the Journal for Research in Mathematics Education. The article served a two-fold purpose. First it gave, in full, the measures devised for examining problem solving processes. These serve as prototypes for straightforward paper-and-pencil measures of the processes involved in problem solving. Such tools are of use both for teachers (to test the efficacy of their instruction) and researchers (as an inexpensive alternative to protocol analysis for some research purposes). Second, the paper documented the results of the problem solving course, in a variety of ways. To put it briefly, the evidence indicates that students given direct instruction in problem solving skills (construed broadly, as characterized below and in the appendices) will show marked improvement in their problem solving performance -- even when working on mathematical problems that are unlike those that they studied in the course.

Some additional experimentation that cut across all three studies provided further evidence of the impact of the course. Research in a variety of domains

(e.g. chess, algebra) indicates that people with substantial practice in a domain have "vocabularies" of familiar objects (words, problems, situations, etc.) to which they have nearly automatic responses; a chess master "sees" a familiar position, for example, and does not have to analyze the full complexity of that position in order to determine his next move. The posession of such "problem schemata" is an important part of routine "expert" problem solving, and the absence of such well developed patterns is, hypothetically, one cause of students' difficulties. Notice that the absence of such patterns for students need not be in new domains; it may simply be that, even with fully accessible problems, students have not yet perceived underlying regularities in solution paths. Of course what one "sees" in a problem statement often determines how one will approach it.

Our research used problems ostensibly accessible to high school students: problems from geometry, elementary combinatorics, algebra, etc. Thus our "novices" (college freshmen) had an adequate mathematical background to deal with these questions. The results of our study indicate substantial differences in what the experts (college faculty) and the novices "see" in the problem statements. Asked to classify "which problems are related in that they would be solved the same way," students classify together problems that share the same "surface structure" -- problems that deal with the same mathematical objects (e.g. circles, functions, or whole numbers). In contrast, experts will often disregard the objects in problem statements to classify together problems that are, on the surface, quite dissimilar: for example, a problem with whole numbers and a problem with functions will be classified together, because both "have implicit negatives in them, so

they're most likely contradiction problems." The elaboration of such structures is a part of both studies 1 and 2. Most dramatically in terms of study 3, our research provided clear evidence that students' perceptions change with experience. As a result of the problem solving course, the students' perceptions came to resemble those of experts (though not perfectly, of course). In contrast, a control group's perceptions of problems remained essentially unchanged. These results, written up as "Problem perception and knowledge structure in expert and novice mathematical problem solvers," (Appendix B) will appear in the Journal of Experimental Psychology. Our early results were also reported elsewhere. A summary report of the underlying theory and the first year's work was delivered at the IV International Congress on Mathematical Education in Berkeley in August, 1980, and will appear ("Toward a testable theory of problem solving," Appendix C) in the Proceedings of the Congress. The research was also discussed in "Recent advances in mathematics education: ideas and implications," (Appendix D), to appear in the Mathematics Education Monograph published by the Mathematics Council of the Alberta Teachers' Association.

The perception studies represented one aspect of the differences between expert and novice performance in problem solving. In this case the differences were the result of familiarity with a domain where the problems were "routine" for the experts. There are other differences, however, that emerge when experts deal with unfamiliar problems. Their success is often a result of what has been called "executive" or "strategic" ability. Roughly, the idea is as follows. In a non-routine problem solving circumstance, there may be any of a large number of plausible paths to a solution. Not only must one have a ready

arsenal of problem solving techniques, but one must be efficient as well. One or two bad "executive" decisions -- pursuing a "wild goose chase" or failing to follow up on a good lead -- can doom a solution to failure, while one or two wise decisions can allow a person who actually "knows less" to solve a problem. A key result of our analyses of videotapes was the development of a framework for the macroscopic analysis of problem solving protocols, focusing on the executive decisions that "make or break" a problem solution. The framework, and the results, presented in "Episodes and executive decisions in mathematical problem solving," (Appendix E, to appear in Acquisition of Mathematics Concepts and Processes, R. Lesh and M. Landau, Eds.) indicate that experts have a well developed set of metacognitive behaviors that keep them "on track," and that the absence of such behaviors may well be a major cause of students' poor problem solving performance.

Issues regarding the study of metacognition, and in general, the interpretation of "verbal data" from protocols such as those gathered in this project are extremely thorny, and not well understood. Part of the project was devoted to clarifying them. A methodological paper,"On the analysis of two-person protocols," (Appendix F, Journal of Mathematical Behavior, in press) elaborated some of the underlying issues as well as presenting the rationale for the framework. This was one of two methodological papers, the second of which was the final study of expert-novice differences in problem solving.

This last paper, "Beyond the purely cognitive: metacognition and social cognition as driving forces in intellectual performance"(Appendix G, Cognitive Science, in press) lays out a broad framework for the analysis of verbal data.

As the title suggests, it indicates that many of the determinants of students' intellectual behaviors lie outside the realm of the "purely cognitive;" that an individual's beliefs about (for example) the way that mathematics is done, what is "legitimate" mathematics, and how mathematics relates to the "real world" may well determine both his choice of approaches to problems and how well he succeeds at them. Our research indicates that even well-trained college students approach mathematics in a surprisingly naive, "purely empirical" way that (1) is antithetical to the approach taken by experienced mathematicians, and (2) serves as a barrier to the development of their mathematical skills. Moreover, it indicates that mathematical instruction that is not compatible with students' belief systems may be doomed to failure, because it "falls upon deaf ears." The paper provides a model of students' beliefs about Euclidean geometry, and derives some consequences of those beliefs. More broadly, it tries to characterize the spectrum of cognitive issues that one must examine (from access to domain-specific facts and procedures to schemata and heuristics, to metacognition, to belief systems) in order to "make sense" of verbal data. These are serious implications for both teaching and research.

In sum, the project has met all of the goals set out for it. At the methodological level, a number of new tools and perspectives have been developed. The measures given in Appendix A provide a straightforward way of capturing a range of problem solving processes on a paper-and-pencil test. The card sort technique developed for measuring problem perception (Appendix B) offered a way of examining changes in students' perceptions as they developed expertise. Appendix E offered a new methodology for parsing protocols and focusing on

"executive" decisions that may "make or break" a solution. Appendix G demonstrated how belief systems impinge upon "purely cognitive" behavior, and that students' behavior can be modeled "from the top down" with an emphasis on such behaviors. Appendices F and G covered a broad spectrum of methodological issues related to the interpretation of verbal data gathered during problem solving sessions. As indicated above, each of these methodologies served to further the goals of the project. They provided evidence of the dimensions along which expert and novice problem solving performance differ, that carry with them implications both for classroom instruction and future research. As study 3 demonstrated, the practical implementation of these ideas can have a strong positive effect on students' problem solving skills.

Dissemination

At this point the major results from the project have all been written up, as indicated in the publication citations and appendices. All the reports have been accepted for publication. As indicated in the proposal, preprints were distributed to a large mailing list including the ONR distribution list. Prior to submission for publication the results in all of the papers were presented at national and international meetings of various scholarly societies (Mathematical Association of America, American Educational Research Association, International Congress on Mathematical Education, International Group for Psychology and Mathematics Education, National Council of Teachers of Mathematics, the Psychonomic Society). During the grant period the PI gave more than two dozen talks and workshops derived directly from the research. Finally, the PI has written about half of a projected book dealing with the research issues covered by the project. The book will, of course, acknowledge NSF support.

Publication Citations

A: Measures of Problem Solving Performance and of Problem Solving Instruction. Journal for Research in Mathematics Education, Vol. 13, No. 1, January 1982.

B: Problem Perception and Knowledge Structure in Expert and Novice Mathematical Problem Solvers, with D. Herrmann. ERIC document E.D. 200-609. Journal of Experimental Psychology: Learning, Memory, and Cognition, in press.

C: Toward a Testable Theory of Problem Solving. In The Proceedings of the IV International Congress on Mathematical Education, in press.

D: Recent Advances in Mathematics Education: Ideas and Implications. In Mathematics Education Monograph of the Mathematics Council of the Alberta Teachers' Association (S. Rachlin, Ed.), in press.

E: Episodes and Executive Decisions in Mathematical Problem Solving. In Acquisition of Mathematics Concepts and Processes (R. Lesh and M. Landau, Eds.) New York: Academic Press, in press. Also ERIC document E.D. 201-505.

F: On the Analysis of Two-Person Problem Solving Protocols. Journal of Mathematical Behavior, in press.

G: Beyond the Purely Cognitive: Metacognition and Social Cognition as Driving Forces in Intellectual Performance. Cognitive Science, in press.

H: Some Thoughts on Problem Solving Research and Mathematics Education. In Perspectives on Research in Mathematical Problem Solving (Frank Lester, Ed.). Philadelphia: Franklin Institute Press, in press.

Scientific Collaborators

A. Research Associate

Douglas J. Herrmann, Associate Professor of Psychology, Hamilton College

B. Grader/Consultant

Andrzej Goralski, University of Warsaw

C. Research Assistant

Johanna P. Zecker, B.A. in Psychology, University of Michigan

D. Student Assistants

Timothy Keiter, Hamilton College Senior Fellow (Wrote Senior Thesis on problem solving.)

Betsy Clayton, Christian Hempel, Robert Kantrowicz, Abed Mougharbel, Michael Stankewicz, Ellen Tourtelot, Hamilton College students.

E. Staff

Mary Lou Millar, secretary

Appendix A was removed because of copyright restrictions. It appears in the "Journal for Research in Mathematics Education," January 1982, Volume 13, number 1, pages 31-49.

PROBLEM PERCEPTION AND KNOWLEDGE STRUCTURE IN EXPERT AND NOVICE MATHEMATICAL PROBLEM SOLVERS

Alan H. Schoenfeld
Graduate School of Education and
Department of Mathematics
The University of Rochester
Rochester, NY 14627

Douglas J. Herrmann
Department of Psychology
Hamilton College
Clinton, NY 13323

SE 038 234

Running Head: Mathematical Problem Perception

This work was supported by a grant from the National Science Foundation, under grant SED 79-19049.

To appear in the Journal of Experimental Psychology: Learning, Memory, & Cognition

Abstract

Evidence regarding the relationships between problem perception and expertise has customarily been obtained indirectly, through contrasting group studies such as expert-novice comparisons. Differences in perception have been attributed to differences in expertise, despite the fact that the groups compared generally differ on a number of other major attributes such as aptitude. This study explored the relationship between perception and proficiency directly.

Students' perceptions of the structure of mathematical problems were studied as the students gained expertise in mathematical problem solving in a month-long intensive problem solving course. Perceptions were measured using a card sorting task, employing cluster analysis and comparing the students' sorting with a sorting done by experts. The data obtained prior to instruction provides direct evidence replicating and extending results from related fields. Experts appear to base their perceptions of problem relatedness upon principles or methods relevant for problem solution, while novices tend to classify problems with the same "surface structure" (i.e. words or objects described in the problem statement) as being highly related.

The data after instruction indicated a strong shift in the students' perceptions, with their post-instruction sorting much more closely approximating those of the experts. These data permit the direct conclusion that criteria for problem perception shift as students' knowledge bases become more richly structured.

Problem Perception and Knowledge Structure in Expert and Novice Mathematical Problem Solvers

Theories of problem solving commonly hold that the mental representation of problems influences how people perceive problems. Moreover, as experience leads to better problem solving, the quality of problem representation is expected to improve with corresponding improvement in problem perception (Chi, Feltovich, and Glaser, in press; Heller and Greeno, 1979; Hayes and Simon, 1976; Newell and Simon, 1972). At one end of the spectrum, the correct perception of a problem may cue access to a "problem schema" which suggests a straight-forward method of solution or a more or less automatic response (Hinsley, Hayes, and Simon, 1978; Chase and Simon, 1973). At the other end, an incorrect perception may send one off on a "wild goose chase." Since problem perception is conceived to be a crucial component of problem solving performance, research on the change in problem perception with the acquisition of expertise has increasingly received more attention (Simon and Simon, 1980; Eylon and Reif, Note 1; Reif, 1979; Larkin, McDermott, Simon, and Simon, 1980).

Early evidence consistent with the hypothesized relationship between expertise and perception was provided in a series of studies by Shavelson (1972, 1974; Shavelson and Stanton, 1975) that indicated that as students learn a discipline, their knowledge of the structural relationships among parts of the discipline become more like that of experts. However, Shavelson's procedures did not directly assess how his subjects perceived problems, and therefore his results do not directly address the perception/expertise hypothesis.

More direct evidence about problem perception and expertise has been

provided by a series of studies in various domains that contrast the problem perceptions of a group of experts in each domain with the perceptions of a group of novices. For example, expert chess players perceive board positions in terms of patterns or broad arrangements, whereas novices do not (de Groot, 1965; Chase and Simon, 1973). Experts in physics perceive problems to be similar if the principles used to solve them, called the "deep structure," coincide. In contrast, novices perceive them as similar if the objects referred to in the problem, or the terms of physics used in the statement, called the "surface structure," coincide (Chi, Feltovich, and Glaser, in press). Two studies on problem perception in mathematics used algebra as their subject domain (Chartoff, 1977; Silver, 1979). There is a consensus regarding the structural isomorphism of algebra word problems, so in both studies problem structures were assigned a priori by the experimenters and no experimental data was collected from experts. In both cases, students who were proficient at solving algebra word problems exhibited a greater degree of agreement with the experimenters' perceptions of the problems than did less proficient students.

The evidence regarding the relationship between expertise and perception, while strong, is indirect. Although expert-novice studies do show that experts and novices differ in problem perception, the design of these studies precludes unequivocal conclusions about the origins of these differences. For example, relative to novices, experts are usually older, more trained, more experienced, and most likely possessed of better aptitude for the subject domain.

Presumably, expert-novice differences in perception are rooted in differences in expertise (training and experience), but they may also be influenced by other psychological properties, for example aptitude. Note that contrasting group designs involving people of the same age may still confound expertise with aptitude. The ambiguous outcome of contrasting group design is, of course,

not unique to studies of expertise and problem perception; the difficulties of the design are well known, and in some areas of psychology these difficulties are regarded as presenting insurmountable obstacles to inference (Schaie, 1977). The present study sought to investigate the effects of expertise on perception in a design that avoids these difficulties: a design that examines problem perceptions in a group of individuals that, with training and experience, improved in problem solving proficiency.

The relationship of perception and expertise was studied in a repeated measures design involving the discipline of college mathematics. Problem perception was assessed before and after training by having students sort a set of math problems. One group of students (hereafter called the experimental group) took a month-long problem-solving course between the sortings. Another group (called the control group) took a month-long course in computer programming between the sortings. In addition, a group of mathematics experts also completed the sort once. This study permits clear assessment of the relationship of problem perception and expertise in the following way. The influence of mathematical training on problem perception may be assessed by comparing the sorting of experimental and control subjects before and after training. If the experimental subjects show sorting after training different than control subjects, inferences about the mathematical improvements in the experimental subjects may be drawn relative to the sorting of the experts. Evidence showing that training affects problem perception and that training fosters problem perception like experts cannot be attributed in this study to differences in individuals (age, maturity, ability and attentional levels). While it is not suggested here that the findings of contrasting groups' studies were not due to differences in expertise, the present procedure provides a clearer assessment of the relationship of problem perception and expertise.

Method

Subjects

Nineteen freshmen and sophomores at Hamilton College, hereafter called novices, participated in the experiment. All of the students had from 1 to 3 semesters of college mathematics prior to the experiment. Eleven of the students (the experimental group) served without pay as a condition of enrollment in a problem solving course, which was the experimental treatment. Eight of the students (the control group) were paid a total of $20 each for participating.

In addition, nine mathematics professors from Hamilton College and Colgate University participated without pay.

Materials

Thirty-two problems were chosen for the study. Each is accessible to students with a high school background in mathematics, dealing with objects familiar from the high school curriculum; none requires calculus for its solution. Each problem was assigned an *a priori* mathematical "deep structure" and a mathematical "surface structure" characterization. The problems used in the study are listed in Appendix A. The characterizations of the problems may be seen in the cluster diagrams (figures 1,2,3).

"Deep structure" refers to the mathematical principles necessary for solution, as identified by the first author who is a mathematician. For example, problems 15 and 17 are both "uniqueness" arguments to be solved by contradiction, although problem 15 deals with geometric objects and problem 17 with functions. These characterizations were independently corroborated by another mathematician. Of the 32 problems, the "deep structure" assessments were literally or essentially agreed upon by this other mathematician for all but three problems (which were perceived in a different but not contradictory fashion). This level of

agreement on deep structure assignments is comparable to that recently reported for physics problems (Chi, et al., in press). "Surface structure" represents a naive characterization of a problem, based on the most prominent mathematical objects that appear in it (polynomials, functions, whole numbers) or the general subject area it comes from (plane or solid geometry, limits). Thus problems 15 and 17 discussed above would be considered a "plane geometry" and a "function" problem respectively.

In addition, two forms of a mathematical problem solving test were used in the study. The tests each had five problems worth 20 points, and were matched for mathematical content. These examinations and a predetermined scheme for awarding partial credit had been pilot tested, with the grading scheme achieving interjudge reliability of greater than .90. Form 1 of the test is given in Appendix B.

Procedures

Both the experimental and control groups performed the card sort and took form 1 of the mathematics test immediately preceding the intensive "winter term" at Hamilton College. Both groups repeated the card sort and took form 2 of the mathematics test a month later, immediately following the conclusion of the winter term. The experts performed the sort once, at their convenience. The sorting procedures were as follows.

Each of the 32 problems was typed on a "3x5" card. Each subject read through the problems in a random order and decided which problems, if any, were "similar mathematically in that they would be solved the same way." A problem that was deemed dissimilar to others was to be placed in a "group" containing one card. Subjects were told that they might return from 1 to 32 "groups" to the experimenter. All subjects finished the task in approximately

20 minutes.

Between the first and second sortings, the experimental treatment consisted of enrollment in a course, "Techniques of Problem Solving," taught by the first author. The class met for two and a half hours per day for 18 days, with daily homework assignments that averaged four to five hours in length. The course focused on general mathematical problem-solving strategies called "heuristics" (Polya, 1945) and stressed a systematic, organized approach to solving problems (Schoenfeld, 1979; 1980). Problems studied in the course were similar to, but not identical to, those used in the sort; Appendix B gives five problems similar to those studied in the course. No mention of problem perception was made during the course. However, students were encouraged to make certain that they had a full understanding of the problem statement before prodeeding with a solution. They were told to examine the conditions of the problem carefully, to look at examples to get a "feel" for the problem, to check for consistency of given data and plausibility of the results, etc. These instructions may well foster the development of improved problem perception.

The control treatment consisted of enrollment in a course, "Structured Programming." The course taught a structured, hierarchical, and orderly way to solve nonmathematical problems using the computer. The students in the course had backgrounds comparable to those of the students in the mathematical problem-solving course, and the course made similar demands in terms of time and effort from the students. Thus this course served as a control for the subject-specific knowledge and skills that might be acquired by the experimental group.

Results

For purposes of comparison with the results of the student sortings, we first present the results for the experts. Figure 1 presents a clustering analysis, using Johnson's (1967) method, of the experts' card sort. Collections of problems exhibiting strong agreement (proximity level exceeding .5, a minimum of 16 out of 32 possible clusters) are bracketed. A brief inspection of figure 1 indicates that the strong clusters are consistently homogeneous

insert figure 1 about here

with regard to deep structure characterizations: in eight of the eleven strong clusters, all of the elements share a common deep structure characterization. In contrast, only four of the eleven strong clusters are homogeneous with regard to surface structure -- and three of these with regard to deep structure as well.

Two measures of the degree of structural homogeneity of figure 1 are given in Table 1. Measure 1 provides, for surface and deep structure respectively, the proportion of strongly clustered pairs that have the same structural representation.[1] Of the twenty-two pairs strongly clustered in figure 1, thirteen (.59) share the same surface structure and eighteen (.82) the same deep structure. We should observe, however, that the surface and deep structures coincided in ten of the twenty-two pairs used in the computation of measure 1. To indicate perceptual preference when the two types of structures conflict, these ten pairs were deleted from the sample for measure 2. With non-coinciding pairs, the proportion of surface-homogeneous pairings for the experts is .25 (3 of 12), and the proportion of deep-homogeneous pairs is .67 (8 of 12).

insert Table 1 about here

Figure 2 presents the cluster diagram of the sorting performed by the combined group of novices (n=19) prior to instruction. In the interest of saving space, the cluster diagrams for the separate experimental and control groups are not given.[2] Inspection of figure 2 indicates a reversal from figure 1, with emphasis on surface structure as the criterion for sorting problems together: eight of ten strong clusters are homogeneous with regard to surface structure, six of ten with regard to deep structure. Of these six, five are also homogeneous with regard to surface structure. The data in table 1 confirm these impressions. Table 1 also provides the data for the separate experimental and control groups prior to instruction. These data, like those for the combined group, indicate that the deep structural relationships between problems were rarely perceived when they ran in contradiction to perceptions of surface structure.

After training, the students who took the problem solving course demonstrated a marked improvement in problem solving performance, while those enrolled in the computer course did not. The mean scores on the mathematics test for the experimental subjects were 21 prior to the course and 73 afterwards. For the control subjects, the mean scores were 14 before and 24 after the course. Analysis of variance on these means showed that scores increased across the term ($F(1,17) = 47.5$, $p<.001$), were greater for experimental rather than for control subjects ($F(1,17) = 130.6$, $p<.001$), and that the increase across the term was not equivalent for experimental and control subjects ($F(1,17) = 48.2$, $p<.001$). Simple effects tests indicated that the term effect was significant for the experimental subjects ($p<.01$) but not for the control subjects. A detailed description of scoring procedures for this measure and of collateral measures may be found in Schoenfeld (1982).

The effect of instruction on problem perception was measured in the ways described above, and also by correlation with the experts' sorting matrix.

Figure 3 presents the cluster analysis of the experimental group's sorting after instruction.

---

insert figure 3 about here

An examination of figure 3 indicates the shift in the students' perceptions. After training, six of eight strong clusters were homogeneous with regard to deep structure, and only four with regard to surface structure; moreover, surface and deep structures coincided in all four of those clusters.

In contrast, the control group's post-instruction sorting shows little change from pre-instruction perceptions. [Again to conserve space, the cluster diagram derived from that sorting, which closely resembles figure 2, is not given. Of ten strong clusters in it, seven are homogeneous with regard to surface structure and only four with regard to deep structure; moreover those four share common deep structures as well.] These results, which indicate a strong change towards "deep structure" perceptions on the part of the experimental group and little or no change on the part of the control group, are given in table 1. Differences between deep and surface proportions were compared across the various conditions with the t approximation to the binomial. Each of the following comparisons (with one exception noted) was significant to at least the ($p<.05$) level, both in direction and size of the differences. Scores within parentheses are reported first for measure 1 (all pairs), then for measure 2 (non-coinciding pairs). The difference for the experts differed, in direction and magnitude, with the difference in the pre-instruction proportions from the experimental group ($t(18) = 2.33$; $t(18) = 4.09$), the control group ($t(15) = 2.61$; $t(15) = 4.98$), and the combined novice group ($t(26) = 1.88$, $p<.1$; $t(26) = 4.81$); also with the post-instruction difference from the control group ($t(15) = 2.31$; $t(15) = 3.99$). Similarly, the differences from

the experimental group after instruction differed ($p<.05$) from the pre-instruction differences from the experimental group ($t(11) = 2.38$; $t(11) = 4.51$), control group ($t(17) = 2.65$; $t(17) = 5.48$), and combined novice group ($t(28) = 2.39$; $t(28) = 5.41$); also from the control group's post-instruction scores ($t(17) = 2.34$; $t(17) = 4.37$).

The comparison of surface and deep structure proportions given above provides an indirect indication that the experimental group's perceptions became more "expert-like" with instruction, while the control group's did not. This relationship was examined more directly by correlating the sorting matrices for each of the treatment groups, before and after instruction, with the sorting matrix obtained from the experts. The correlations are given in table 2. With df = 496, all correlations are significant. The pretest correlations and the control post-test correlations are significantly less ($p<.01$) than the experimental group's post-test correlation.

---

insert table 2 about here

Discussion

The design of this study allows for the direct attribution of the students' changes in problem perception to changes in their problem solving proficiency. This attribution cannot be made unequivocally from any of the contrasting group studies conducted to date, for example the standard expert-novice studies. Note that professors or advanced graduate students in a discipline differ from lower division undergraduates in maturity, cohort group, comfort in testing situations, and most notably, aptitude. A clear understanding of how novices' performance improves in a discipline cannot be obtained by comparing them to a group of experts whose aptitude for the discipline is, in all

likelihood, far beyond that of the novices. Similarly, an understanding of expert perception cannot be obtained by taking as the starting point of that development people whose performance alone makes it unlikely that they will ever be expert in that domain. One might obtain experimental confirmation of the relationship between perception and expertise in contrasting group designs in which the groups had been matched on all variables except expertise (a difficult proposition, and a condition not present in any expert-novice studies with which we are familiar). However, the most direct way to ascertain that relationship is with a repeated measures (longitudinal) design like the one used here.

Two other points should be considered before the specifics of the data are elaborated. First, the nature of deep structure in mathematics is different from that of other domains. For example, elementary physics is strongly principle-driven, and the subject matter is organized and taught according to those principles. Mathematics is not organized and taught that way, however. One talks about methods of solution, rather than principles; and the curriculum is organized around topics rather than around those methods, which are simply the tools used to solve them. Thus, there does not exist an a priori consensus about the structure of the problems used in this study that would lead one to predict with confidence the particular pattern of results repeated in figure 1. The absence of such a consensus makes the consistency of the present results more impressive. The word "novice" in this study does not mean rank beginner; the students in this study had extensive mathematical backgrounds and were, in the sort, reading problems accessible to them. The surface labels reflect this, for example in the labels for problems 2 and 11. Surely, one would be surprised if college students could not see that integer combinations of weights and integer combinations of costs called for the same mathematics! (This would

not necessarily be the case with fifth graders, for example).

The data in table 1 provide a strong indication that the experimental group's perceptions of problem structure shifted from a basis in surface structure to a basis in deep structure. An examination of the experimental group's post-instruction cluster {9, 17, 10} illustrates the change in problem perception. Problem 9 deals with whole numbers and, prior to instruction, was sorted with two other "whole number" problems in a homogeneous surface structure cluster. Problem 17 deals with abstract functions, and, prior to instruction, was (barely) clustered with a problem that presented a very complex polynomial function for analysis. Problem 10 deals with polynomials, and was placed in a strong cluster all three of whose terms had the surface label "polynomials, roots." Each of these problems is solved by the mathematical technique known as proof by contradiction and, despite their differing surface characterizations, they are all placed in the same cluster after instruction. The broad shift towards expert perceptions is confirmed in figure 2, which shows the correlation between experts' and the experimental group's sorting matrices jumping from .540 (before instruction) to .723 (after instruction), the only significant ($p<.01$)change in correlation. This rather dramatic shift, after a short period of time, indicates that instructional treatments that focus on understanding and performance can have a strong impact on perceptions.

Despite the strong shift in the students' sort, the experimental group's performance after instruction cannot be truly called "expert-like." The experts' extended knowledge and experience allow them perceptions inaccessible to the novices. Consider, for example, the three bracketed clusters including problem 1: novice {1, 32, 9}; experimental {1, 3, 21}; expert {1, 3}. The experimental group drops problems 32 and 9, which are similar to problem 1 only in that they deal with whole numbers. Problem 3, which shares the same deep structure

as problem 1, is added. The mimicry of expert perceptions is not exact, however: problem 21 is added as well. The addition of problem 21 provides an indication of the "intermediate" status of the experimental group. Problems 12 and 21 were included in the card sort to see if the experts would cluster them together. Underlying the experts' perception of problem 21 is the observation that multiples of 9 and multiples of 4 both include multiples of 36 (their intersection), and that one must compensate for subtracting the first two sets by adding the third. This is structurally similar to the rule $N(A \cup B) = N(A)+N(B) -N(A \cap B)$ upon which problem 12 is based. This is a rather subtle observation. While experts' experience with combinatorics problems might make such an observation readily accessible, novices even with training cannot be expected to see such subtleties. In the absence of such knowledge, it is plausible to think that "looking for patterns" will help to solve problem 21 -- and thus to sort it with two other "patterns" problems.

The research described here supports and extends previous research on problem perception. The novices' card sort indicated that, in the broad domain of general mathematical problem solving, students with similar backgrounds will perceive problems in similar ways. This is consistent with previous research in mathematics, which had considered only word problems in algebra (Hinsley, et al., 1977; Chartoff, 1977; Silver, 1979). Like research in physics (Chi, et al., in press), it suggests that surface structure is a primary criterion used by novices in determining problem relatedness. Moreover, it verifies directly that students' problem perceptions change as the students acquire problem-solving expertise. Not only their performance, but their perceptions, become more like experts'.

In general, questions regarding the deep structures in individual disciplines and the nature of experts' perceptions in those disciplines are

more complex than those regarding surface structures and novices' perceptions in them. The differences between the structures of mathematics and physics were discussed above. In another discipline, research on chess perception (de Groot, 1965; Chase and Simon, 1973) indicates that experts' perceptions of routine problems (similar in a way to the routine physics and mathematics problems discussed above) may be based on the acquisition of a "vocabulary" of known situations which is not necessarily principle-based. Further research might profitably be directed towards the elucidation of how deep structures differ across disciplines and how problem perceptions evolve with the acquisition of expertise in different domains.

Reference Note

1. Eylon, B., & Reif, F. Effects of internal knowledge organization on task performance. Expanded version of a paper presented at the annual meeting of the American Educational Research Association, Berkeley, April 1979.

References

Chartoff, B. T. An exploratory investigation utilizing a multidimensional scaling procedure to discover classification criteria for algebra word problems used by students in grades 7-13 (Doctoral dissertation, Northwestern University, 1976). Dissertation Abstracts International, 1977, 37, 7006A.

Chi, M., Feltovich, P. J., & Glaser, R. Representation of physics knowledge by experts and novices. Cognitive Science, in press.

Chase, W. G. & Simon, H. A. Perception in chess. Cognitive Psychology, 1973, 4, 55-81.

de Groot, A. Perception and memory versus thought: some old ideas and recent findings. In B. Kleinmunz (Ed.), Problem Solving. New York: Wiley, 1966.

Hinsley, D. A., Hayes, J. R. & Simon, H. A. From words to equations: meaning and representation in algebra word problems. In P. A. Carpenter & M. A. Just (Eds.), Cognitive processes in comprehension. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32 241-54.

Larkin, J., McDermott, J., Simon, D. & Simon, H. A. Expert and novice performance in solving physics problems. Science, 1980, 208, 1335-1342.

Schaie, K. W. Quasi-experimental research designs in the psychology of aging. In J. E. Birren, and K. W. Schaie (Eds.), Handbook of the Psychology of Aging. Van Nostrand Reinhold: New York, 1977.

Schoenfeld, A. H. Can heuristics be taught? In J. Lochhead & J. Clement (Eds.), Cognitive process instruction. Philadelphia, Pa.: Franklin Institute Press, 1979.

_____. Teaching problem solving skills. American Mathematical Monthly, 1980. 87(10).

_____. Measures of problem solving performance and of problem solving instruction. Journal for Research in Mathematics Education, Jan. 1982, 13, No. 1.

Shavelson, R.J. Some aspects of the correspondence between content structure and cognitive structure in physics instruction. Journal of Educational Psychology, 63, No. 3, 225-234.

_____. Methods for examining representations of a subject-matter structure in a student's memory. Journal of Research in Science Teaching, 11, No. 3, 231-249.

_____, and Stanton, G.C. Construct validation: methodology and application of three measures of cognitive structure. Journal of Educational Measurement, 12, No. 2, 67-85.

Silver, E.A. Student perceptions of relatedness among mathematical verbal problems. Journal for Research in Mathematics Education, 1979, 10(3), 195-210.

Footnotes

The authors thank Chris Hempel, Mike Stankiewicz, Rob Kantrowitz, and Dorcque Schoenfeld for help on this research. This material is based upon work supported by the National Science Foundation at Hamilton College under Grant No. SED-79 19049. Requests for reprints should be sent to Alan H. Schoenfeld, Graduate School of Education and Department of Mathematics, The University of Rochester, Rochester, N.Y. 14627.

[1] We wish to thank Jim Greeno and Alice Healy for suggesting the measure and stengthening the discussion.

[2] All three diagrams are quite similar. The matrix from which Figure 2 was derived was strongly correlated with both the experimental pretest matrix ($r = .918$, $df = 498$, $p<.001$) and the control pretest matrix ($r = .889$, $df = 496$, $p<.001$).

Table 1

The proportion of strongly clustered pairs in which both problems share the same representation

(Number of pairs given in parentheses)

	Measure 1 (All pairs)		Measure 2 (Non-coinciding pairs)	
	Surface Structure	Deep Structure	Surface Structure	Deep Structure
Experts	.59(22)	.82(22)	.25(12)	.67(12)
Experimental, Pretest	.81(26)	.58(26)	.58(12)	.08(12)
Control, Pretest	.91(23)	.57(23)	.82(11)	.09(11)
Combined, Pretest	.76(21)	.62(21)	.67(9)	.11(9)
Experimental, Posttest	.58(24)	.79(24)	.09(11)	.55(11)
Control, Posttest	.83(24)	.58(24)	.64(11)	.09(11)

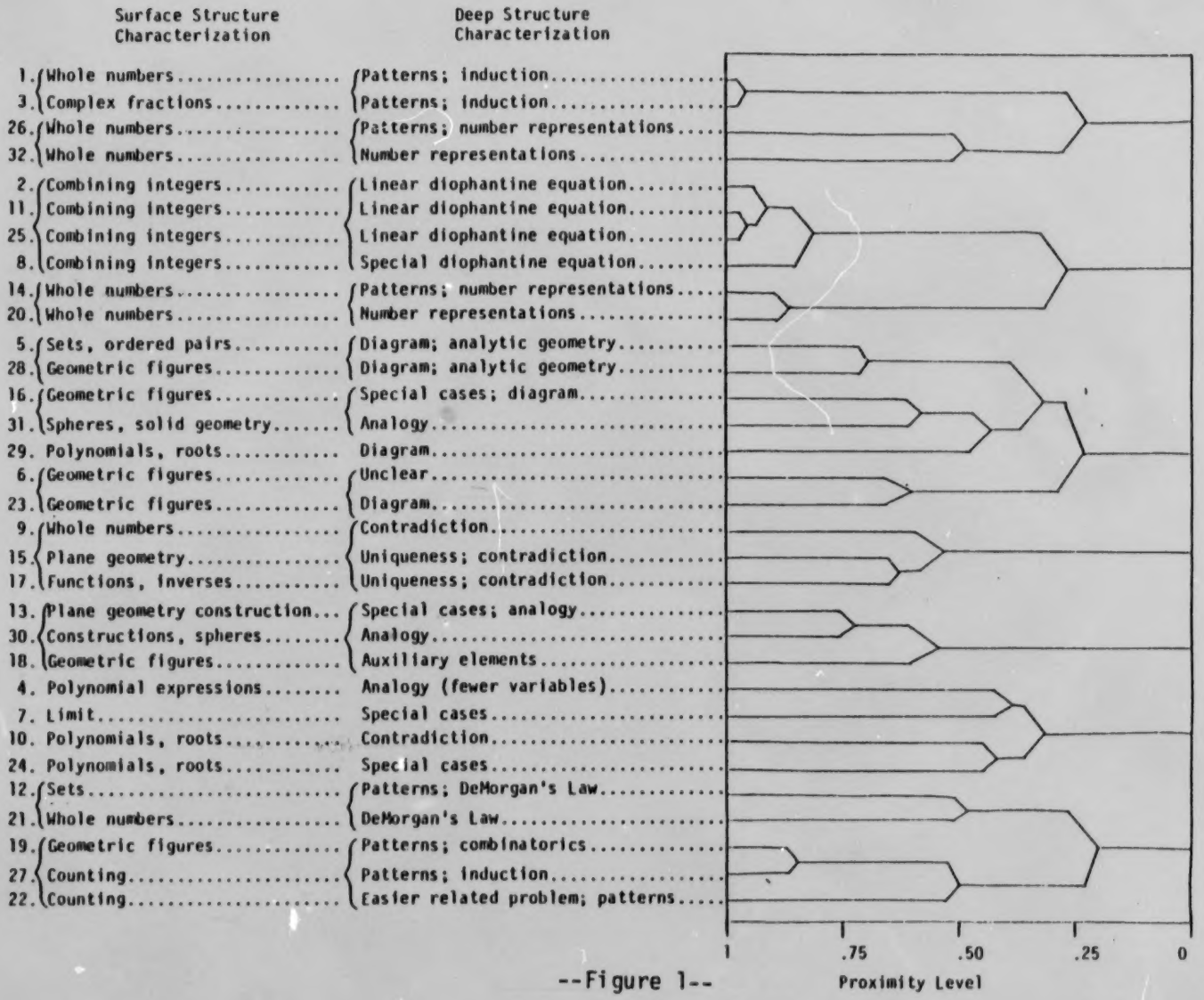
Table 2

Correlations Between Sorting Matrices

of Novices (Given at Left) with Expert Sort

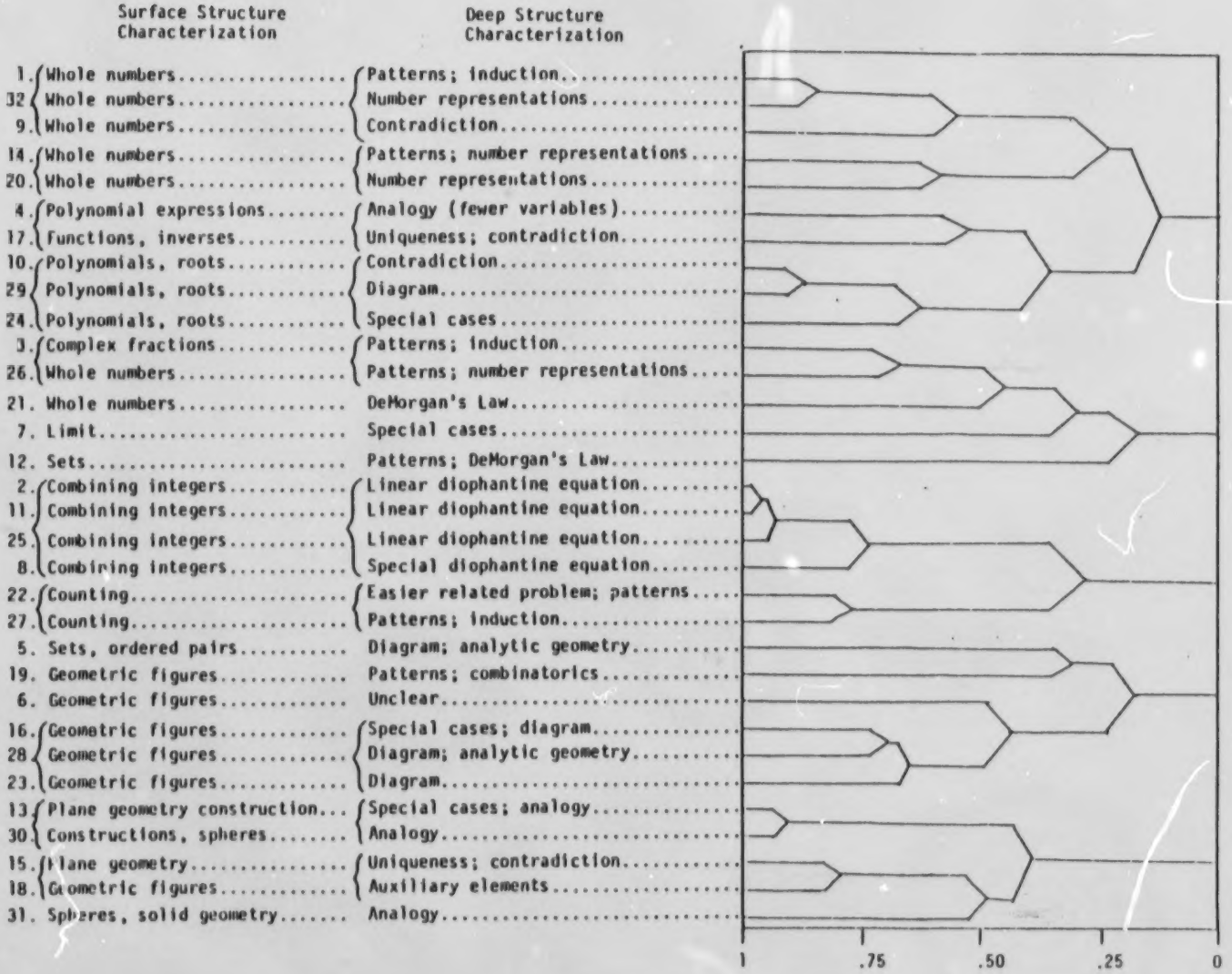
Control, Pretest	.551
Experimental, Pretest	.540
Combined, Pretest	.602
Control, Post Test	.423
Experimental, Post Test	.723

Cluster Analysis of Experts' Card Sort


Surface Structure Characterization
Deep Structure Characterization
1. Whole numbers ... Patterns; induction
3. Complex fractions ... Patterns; induction
26. Whole numbers ... Patterns; number representations
32. Whole numbers ... Number representations
2. Combining integers ... Linear diophantine equation
11. Combining integers ... Linear diophantine equation
25. Combining integers ... Linear diophantine equation
8. Combining integers ... Special diophantine equation
14. Whole numbers ... Patterns; number representations
20. Whole numbers ... Number representations
5. Sets, ordered pairs ... Diagram; analytic geometry
28. Geometric figures ... Diagram; analytic geometry
16. Geometric figures ... Special cases; diagram
31. Spheres, solid geometry ... Analogy
29. Polynomials, roots ... Diagram
6. Geometric figures ... Unclear
23. Geometric figures ... Diagram
9. Whole numbers ... Contradiction
15. Plane geometry ... Uniqueness; contradiction
17. Functions, inverses ... Uniqueness; contradiction
13. Plane geometry construction ... Special cases; analogy
30. Constructions, spheres ... Analogy
18. Geometric figures ... Auxiliary elements
4. Polynomial expressions ... Analogy (fewer variables)
7. Limit ... Special cases
10. Polynomials, roots ... Contradiction
24. Polynomials, roots ... Special cases
12. Sets ... Patterns; DeMorgan's Law
21. Whole numbers ... DeMorgan's Law
19. Geometric figures ... Patterns; combinatorics
27. Counting ... Patterns; induction
22. Counting ... Easier related problem; patterns
1
.75
.50
.25
0
Proximity Level

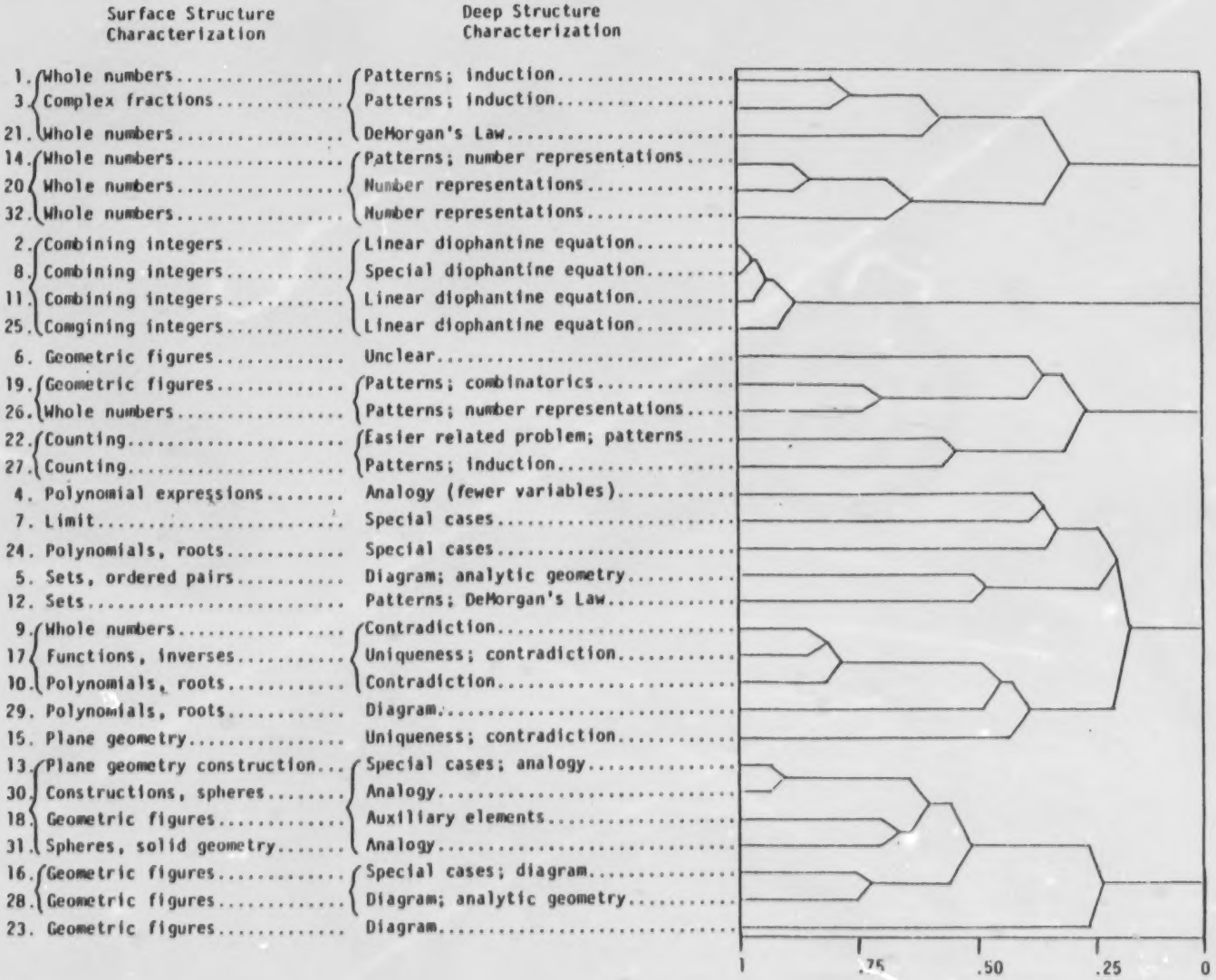
--Figure 1--

Cluster Analysis of Combined Novices' Card Sort


Surface Structure Characterization
Deep Structure Characterization
1. Whole numbers — Patterns; induction
32 Whole numbers — Number representations
9. Whole numbers — Contradiction
14. Whole numbers — Patterns; number representations
20. Whole numbers — Number representations
4. Polynomial expressions — Analogy (fewer variables)
17. Functions, inverses — Uniqueness; contradiction
10. Polynomials, roots — Contradiction
29 Polynomials, roots — Diagram
24. Polynomials, roots — Special cases
3. Complex fractions — Patterns; induction
26. Whole numbers — Patterns; number representations
21. Whole numbers — DeMorgan's Law
7. Limit — Special cases
12. Sets — Patterns; DeMorgan's Law
2. Combining integers — Linear diophantine equation
11. Combining integers — Linear diophantine equation
25. Combining integers — Linear diophantine equation
8. Combining integers — Special diophantine equation
22. Counting — Easier related problem; patterns
27. Counting — Patterns; induction
5. Sets, ordered pairs — Diagram; analytic geometry
19. Geometric figures — Patterns; combinatorics
6. Geometric figures — Unclear
16. Geometric figures — Special cases; diagram
28 Geometric figures — Diagram; analytic geometry
23. Geometric figures — Diagram
13 Plane geometry construction... — Special cases; analogy
30. Constructions, spheres — Analogy
15. Plane geometry — Uniqueness; contradiction
18. Geometric figures — Auxiliary elements
31. Spheres, solid geometry — Analogy
1
.75
.50
.25
0
Proximity Level

--Figure 2--

Cluster Analysis of Experimental Group's Card Sort, After Instruction


Surface Structure Characterization
Deep Structure Characterization
1. Whole numbers ... Patterns; induction
3. Complex fractions ... Patterns; induction
21. Whole numbers ... DeMorgan's Law
14. Whole numbers ... Patterns; number representations
20 Whole numbers ... Number representations
32. Whole numbers ... Number representations
2. Combining integers ... Linear diophantine equation
8. Combining integers ... Special diophantine equation
11. Combining integers ... Linear diophantine equation
25. Comgining integers ... Linear diophantine equation
6. Geometric figures ... Unclear
19. Geometric figures ... Patterns; combinatorics
26. Whole numbers ... Patterns; number representations
22. Counting ... Easier related problem; patterns
27. Counting ... Patterns; induction
4. Polynomial expressions ... Analogy (fewer variables)
7. Limit ... Special cases
24. Polynomials, roots ... Special cases
5. Sets, ordered pairs ... Diagram; analytic geometry
12. Sets ... Patterns; DeMorgan's Law
9. Whole numbers ... Contradiction
17 Functions, inverses ... Uniqueness; contradiction
10. Polynomials, roots ... Contradiction
29. Polynomials, roots ... Diagram
15. Plane geometry ... Uniqueness; contradiction
13. Plane geometry construction ... Special cases; analogy
30. Constructions, spheres ... Analogy
18. Geometric figures ... Auxiliary elements
31. Spheres, solid geometry ... Analogy
16. Geometric figures ... Special cases; diagram
28. Geometric figures ... Diagram; analytic geometry
23. Geometric figures ... Diagram
1
.75
.50
.25
0
Proximity Level

--Figure 3--

APPENDIX A

PROBLEMS USED IN CARD SORT

1. Show that the sum of consecutive odd numbers, starting with 1, is always a square. For example,

$$1+3+5+7 = 16 = 4^2.$$

2. You have an unlimited supply of 7 pound weights, 11 pound weights, and a potato which weighs 5 pounds. Can you weigh the potato on a balance scale? A 9 pound potato?

3. Find and verify the sum

$$\frac{1}{1.2} + \frac{2}{1.2.3} + \frac{3}{1.2.3.4} + \ldots + \frac{n}{1.2.3\ldots(n+1)}$$

4. Show that if x, y, and z are greater than 0,

$$\frac{(x^2+1)(y^2+1)(z^2+1)}{xyz} > 8.$$

5. Find the smallest positive number m such that the intersection of the set of all points $\{(x,mx)\}$ in the plane, with the set of all points at distance 3 from (0,6), is non-empty.

6. The lengths of the sides of a triangle form an arithmetic progression with difference d. (That is, the sides are a, a+d, a+2d.) The area of the triangle is t. Find the sides and angles of this triangle. In particular, solve this problem for the case d = 1 and t = 6.

7. Given positive numbers a and b, what is

$$\lim_{n\to\infty} (a^n + b^n)^{1/n}?$$

8. In a game of "simplified football," a team can score 3 points for a field goal and 7 points for a touchdown. Notice a team can score 7 but not 8 points. What is the largest score a team <u>cannot</u> have?

9. Let n be a given whole number. Prove that if the number (2^n-1) is a prime, then n is also a prime number.

10. Prove that there are no real solutions to the equation

$$x^{10} + x^8 + x^6 + x^4 + x^2 + 1 = 0$$

11. If Czech. currency consists of coins valued 13 cents and 17 cents, can you buy a 20-cent newspaper and receive exact change?

12. If N(A) means "The number of elements in A," then $N(A \cup B) = N(A) + N(B) - N(A \cap B)$. Find a formula for $N(A \cup B \cup C)$.

13. Construct, using straightedge and compass, a line tangent to two given circles.

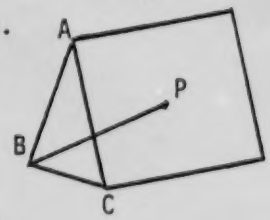
14. Take any odd number; square it; divide by 8. Can the remainder be 3? or 7?

15. You are given the following assumptions:
    i) Parallel lines do not intersect; non-parallel lines intersect.
    ii) Any two points P and Q in the plane determine an unique line which passes between them.
    Prove: Any two distinct non-parallel lines L_1 and L_2 must intersect in an unique point P.

16. Two squares "s" on a side overlap, with the corner of one on the center of the other. What is the maximum area of possible overlap?

17. Show that if a function has an inverse, it has only one.

18. Let P be the center of the square constructed on the hypotenuse AC of the right triangle ABC. Prove that BP bisects angle ABC. [see figure at right.]



19. How many straight lines can be drawn through 37 points in the plane, if no 3 of them lie on any one straight line?

20. If you add any 5 consecutive whole numbers, must the result have a factor of 5?

21. What is the sum of all numbers from 1 to 200, which are not multiples of 4 and 9? You may use the fact that
$$(1+2+\ldots+n) = 1/2\ (n)(n+1)$$

22. Your goal is to convert figure 1 to figure 2. You may move only one disk at a time from one spike to another, and you may never put a larger disk on top of a smaller one. How to?

fig.1 fig.2

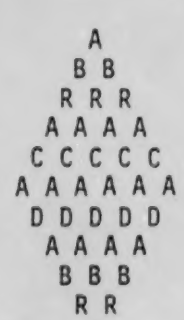
23. Determine the area of a triangle whose sides are given as 25, 50, and 75.

24. If P(x) and Q(x) have "reversed" coefficients, for example
$$P(x) = x^5+3x^4+9x^3+11x^2+6x+2,$$
$$Q(x) = 2x^5+6x^4+11x^3+9x^2+3x+1,$$
What can you say about the roots of P(x) and Q(x)?

25. You have 2 unmarked jugs, one whose capacity you know to be 5 quarts, the other 7 quarts. You walk down to the river and hope to come back with precisely 1 quart of water. Can you do it?

26. What is the last digit of $(\ldots((7^7)^7)^7\ldots)^7$, where the 7th power is taken 1,000 times?

27. Consider the magical configuration shown at right. In how many ways can you read the word "ABRACADABRA?"



28. A circular table rests in a corner, touching both walls of a room. A point on the rim of the table is eight inches from one wall, nine from the other. Find the diameter of the table.

29. Let a and b be given real numbers. Suppose that for all <u>positive</u> values of c, the roots of the equation

$$ax^2+bx+c = 0$$

are <u>both</u> real, positive numbers. Present an argument to show that a must equal zero.

30. Describe how to construct a sphere which circumscribes a tetrahedron (the 4 corners of the pyramid touch the sphere.)

31. Let S be a sphere of radius 1, A an arc of length less that 2 whose endpoints are on the boundary of S. (The interior of A can be in the interior of S.) Show there is a hemisphere H which does not intersect A.

32. Show that a number is divisible by 9 if and only if the sum if its digits is divisible by 9. For example, consider 12345678: 1+2+3+4+5+6+7+8 = 36 = 4x9, so 12345678 is divisible by 9.

Appendix B

Mathematics Test Form 1

1. If S is any set, we define O(S) to be the number of subsets of S which contain an odd number of elements. For example: the "odd" subsets of {A, B, C} are {A}, {B}, {C}, and {A, B, C}; thus O({A,B,C}) = 4. Determine O(S) if S is a set of 26 objects.

2. Suppose you are given the positive numbers p,q,r, and s. Prove that $$\frac{(p^2+1)(q^2+1)(r^2+1)(s^2+1)}{pqrs} \geq 16.$$

3. Suppose T is the triangle given in figure 1. Give a mathematical argument to demonstrate that there is a square, S, such that the 4 corners of S lie on the sides of T, as in figure 2.

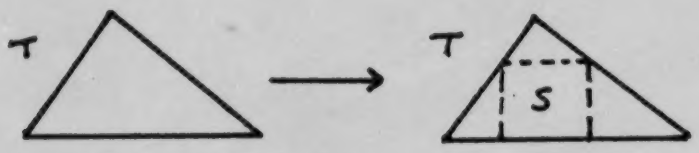


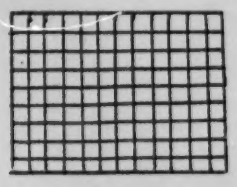
fig. 1 fig. 2

4. Consider the set of equations

$$\left\{\begin{array}{l} ax + y = a^2 \\ x + ay = 1 \end{array}\right\}$$

For what values of "a" does this system fail to have solutions, and for what values of "a" are there infinitely many solutions?

5. Let G be a (9 x 12) rectangular grid, as illustrated to the right. How many different rectangles can be drawn on G, if the sides of the rectangles must be grid lines? (Squares are included, as are rectangles whose sides are on the boundaries of G.)



TOWARD A TESTABLE THEORY OF PROBLEM SOLVING*

Alan H. Schoenfeld
Mathematics Department
Hamilton College
Clinton, N. Y. 13323

*This paper was presented at the 4th International Congress on Mathematical Education, August 1980, Berkeley, California. It provides a brief introduction and overview of my work in problem solving.

Parts of the work described here are supported under NSF Research in Science Education Grant SED 79-19049.

SE038234

Toward A Testable Theory of Problem Solving

The four main issues I address in this paper are the following.

1. What, beyond basic subject matter mastery, serves to explain "expert" mathematical problem solving behavior?
2. What traits do students lack, or what inappropriate traits do they have, which prevent them from approaching problems with the flexibility and resourcefulness of experts?
3. Can we teach students to "solve problems like experts"--and how?
4. What clear, scientific evidence can we offer to support our opinions regarding the first three questions?

The students I refer to will be advanced high school or lower division undergraduate students and the problems "nonroutine," of the type discussed by Polya in Mathematical Discovery. "Experts" can be defined (for the sake of simplicity, and not uniquely) as college mathematics faculty.

The answers from the mathematics education community to the first three questions would, I suspect, involve the word revived by the Honorary President of this Congress, George Polya: "heuristics." The fourth question is harder. There has been little conclusive evidence to date that heuristics "work"--in the sense that students can learn to use them, and improve their problem solving performance thereby. In fact, heuristics are for the most part ignored, dismissed or disdained outside the math-ed community. Herbert Simon, writing to "christen" the new domain of cognitive science, spoke of "cognitive

psychologists, researchers in artificial intelligence, philosophers, linguists, and others who seek to understand the workings of the human mind."[1] Allen Newell, coauthor with Simon of Human Problem Solving, wrote that we are working in the wrong direction: "If we ask what evidence we have that Polya is right . . . the answer is that there is none of a scientific kind. We are all impressed and pleased, that's all."[2] For its own part the math-ed community ignores with equal impugnity the advances made in cognitive science: the 1980 NCTM Yearbook, Problem Solving in School Mathematics, would essentially be unchanged if all the fields listed above by Simon did not exist. The result is a loss to both schools. The interplay between them can, and should be fruitful. I shall discuss here some adaptations of ideas and techniques from cognitive science to examine problem solving via heuristics--and to provide some of the evidence Newell asks for.

We shall outline the framework of a theory. In brief, we argue that there are (at least) three components which are essential for competent problem solving performance in any nontrivial domain:

I. An adequate knowledge base, including access to basic facts, relations, and procedures.

II. The mastery of relevant problem solving techniques--in the case of nonroutine problem solving like that discussed here, the mastery of certain heuristics; and

III. An efficient means of selecting appropriate techniques for application, and in general for using efficiently those resources which the problem solver has at his or her disposal. We shall call this "efficiency expert" a managerial strategy.

I. The Knowledge Base

The first observation we make is that an adequate characterization of the knowledge base for problem solving is more important than it might at first appear. Of course, the problem solver must have the "basic facts" at his disposal. But recent work in cognitive science stresses the difference between factual knowledge and procedural knowledge. The latter includes a knowledge of the conditions under which a particular procedure may or may not be "legal," to what arguments it applies, and so on. Note that the organization of the knowledge base is important: in addition to "knowing" something (that is, being able to discuss it when asked about it) one must know when it is relevant to a particular problem. Otherwise one's knowledge is wasted. An experiment in the psychology of physics learning shows that the organization of one's knowledge base has a strong effect on one's success in solving problems.[3] A result I obtained recently indicates that experts and novices actually "see" different things in problem structures: the criteria which experts use for judging whether two mathematical problems are related are quite different from those used by novices.[4]

The experiment was conducted as follows. Each person was given a collection of 32 problems, and asked to sort the problems into anywhere from 1 to 32 piles, each pile containing problems which were "mathematically related, or would be solved the same way." A computer analysis of the cumulative sorting (the HICLUS program), revealed that the (apparent) criteria for sorting were quite different. Novices clustered problems by what we call their "surface structure:" that is, by the objects which the problems deal with. For example, three problems dealing with the roots of polynomials would be called "related" by the novices, even if one was most appropriately

solved by graphing, another by examining special cases, and the third by contradiction. In contrast, the "expert" clusters were sorted by what we call "deep structure." Problems amenable to an approach by mathematical induction were clustered together, even if one dealt with points and lines, a second with the last digit of a complicated numerical expression, and a third with the coefficients of a polynomial. These perceptions, of course, affect problem solving performance. We are just beginning to deal with the complexity of knowledge structures, and there is much to discover.

II. The Heuristics

The second major component for nonroutine problem solving, as we discussed it above, is the ability to use certain heuristics. Most attempts to document the role of heuristics in problem solving have yielded very equivocal results. This is not surprising if one considers (1) the quite complex and usually underestimated web of skills needed to correctly employ individual heuristics, and (2) our hypothesis that the heuristics alone are not sufficient to guarantee improved problem solving performance. Our treatment will be brief here: we will mention only two studies designed to "tease out" the role of heuristics. See the 1980 NCTM Yearbook and the NCTM's Research in Mathematics Education for extensive discussions of the literature. The first study[5] was designed to see whether students will intuit problem solving heuristics simply by working problems. Under controlled laboratory conditions, two groups of students were trained for identical periods of time on identical problems, with only one group given the explicit heuristics underlying the solutions they were shown. There was a significant difference in performance; the "control" students were not able to use their problem solving experience to solve related problems, while the experimental group was explicitly using

the heuristics. However, the journey from the laboratory to the classroom is a long one.

The second study took place in the classroom. It had two goals: (1) to define some useful and replicable measures, so that other teachers and researchers could replicate the results; (2) to use those measures to verify a substantial improvement in students' problem solving performance. In that study,[4] students were both taught heuristics and a managerial strategy (of sorts); thus the effects of heuristics (or the managerial strategy) alone are hard to sort out. But there was much greater heuristic fluency, correlating with dramatically improved problem solving.

III. Managerial Strategies

Finally, we come to the presence of an efficient "manager" itself. This is also difficult to sort out, for all problem solvers obviously have some managerial abilities. Our first argument is by analogy. In an early study,[6] students who had learned the techniques of integration were divided into two groups. One group studied the "usual" way, each working problems for an average of nine hours. The other group was given a strategy which helped them to select the appropriate techniques. The experimental group averaged seven hours study time, and significantly outperformed the control group. Now, the argument to be made here is that, even in a simple domain, students who lack an efficient manager squander some of their resources. In general, where there are many more choices and many more opportunities to go wrong, the absence of an efficient manager can be debilitating--even if one has the appropriate heuristic abilities.[7]

We are currently developing a scheme for analyzing transcripts of problem solving sessions which focuses on managerial actions. Although we

have only preliminary results, we believe that this scheme may allow us to (1) characterize some "expert" managerial actions which account for efficient problem solving, (2) demonstrate the consequences of poor managerial actions in students' problem solving, and (3) correlate improved performance in problem solving with both heuristic and managerial improvement. If all goes well, the synthesis of ideas from the heuristic school with the techniques from cognitive science will help us to better understand, and teach, problem solving.

References

1. Simon, Herbert. Cognitive Science: The newest science of the artificial. CIP working paper. Department of Psychology, Carnegie-Mellon University, 1979.

2. Newell, Allen. Personal Communication. January 1979.

3. Reif, Fred. Cognitive Mechanisms Facilitating Human Problem Solving in a Realistic Domain: The Example of Physics. Physics Department, University of California, Berkeley, 1979.

4. Schoenfeld, Alan. Measures of students' problem solving performance and of problem solving instruction. Mathematics Department, Hamilton College, 1980.

5. _____. Explicit Heuristic Training as a Variable in Problem Solving Performance. Journal for Research in Mathematics Education, May 1979.

6. _____. Presenting a strategy for indefinite integration. American Mathematical Monthly, October 1978.

7. _____. Teaching Problem Solving Skills. American Mathematical Monthly, in press.

Recent Advances in Mathematics Education:

Ideas and Implications*

Alan H. Schoenfeld
Graduate School of Education and
Department of Mathematics
University of Rochester
Rochester, NY 14627

To appear in the Alberta Teachers' Association Monograph on Problem Solving. Based on a paper presented at the A.M.S. Special Session on Mathematics Education, Davis, CA, 1980. Parts of the research described here were supported by the NSF Research in Science Education grant SED 79-19049

SE 038 234

Recent Advances in Mathematics Education:
Ideas and Implications

There have been major changes in mathematics education research over the past decade. Research in education is now highly interdisciplinary, with contributions from cognitive psychologists, workers in artificial intelligence, etc. There are new people, new perspectives, new methodologies -- and most important, new results. Taken as a whole, these results promise to re-shape our understanding of the learning and teaching processes. In this paper I will discuss one aspect of recent work, and its implications.

The three examples I'm going to discuss in this paper seem on the surface to have little to do with each other. John Seely Brown and Richard R. Burton have done a detailed analysis of the way elementary school children perform certain simple arithmetic operations. John Clement, Jack Lochhead, and Elliot Soloway have studied the way that people translate sentences like "There are six times as many students as professors at this college" into mathematical symbolism. My work consists of an attempt to model "expert" mathematical problem solving, and to teach college freshmen to "solve problems like experts." Yet all three of these studies share a common premise, and their results tend to substantiate it. That premise is the following:

> There is a remarkable degree of consistency in both correct and incorrect mathematical behavior on the part of both experts and novices. This consistency is so strong that it may often be possible to model or simulate that behavior, at a very substantive level of detail.

The implications of this assumption for both the teaching and learning processes are enormous. First, consider the notion that much of our student's incorrect behavior can be simulated--and hence predicted. This means that many of their mistakes are not random, as we often assume, but the result of a consistently applied and incorrectly understood procedure. In consequence, the student does not need to be "told the right procedure"; he needs to be "debugged." This idea lies at the heart of the Brown and Burton work. It is also central to Lochhead and Clement's work, where we will see that the simple process of translating a sentence into algebraic symbols is far more complex than it at first appears. The other side of the coin has to do with the consistency of expert behavior. That, of course, is the assumption made in artificial intelligence--where the attempt is made to model expert behavior in enough detail so that it can be simulated on a computer. If that seems plausible, then another step should seem equally plausible: model expert behavior so that humans, rather than machines, can simulate it. That is, teach students to "solve problems like experts" by training them to follow a detailed model of expert problem solving. That is the idea behind my own work.

Needless to say, a discussion of these three projects barely scratches the surface of what is happening in mathematics education today. I could not hope to be comprehensive in a brief paper like this, and I think it would be a waste of time to simply provide a list of people's names with one or two sentence descriptions of what they are doing. Instead, I have chosen to look at just one idea and to discuss in some detail how three different research projects explore aspects of it. In doing so I hope to at least convey the

flavor of some modern work--and leave you with a taste for more. The bibliography suggests further readings.

2. A Close Look at Arithmetic.

In this section I offer a distillation of Brown and Burton's paper "Diagnostic Models for Procedural Bugs in Basic Mathematical Skills." There is much more in that paper than I can summarize here, and it is well worth reading in its entirety; so are the other papers by them listed in the bibliography. They are now at Xerox PARC, 3333 Coyote Hill Rd., Palo Alto, Ca., 94304. Address reprint requests there.

The key word in the title of their paper is "bug." It is, of course, borrowed from programming terminology--and is fully intended to have all of the connations that it usually does. While a seriously flawed program may fail to run, a program with only one or two minor bugs may run all the time. It may even produce correct answers most of the time. Only under certain circumstances will it produce the wrong answer--and then it will produce that wrong answer consistently.

Often one discovers a bug in a computer program when it produces the wrong answer on a test computation. Now, one might hope to find the bug by reading over the listing of the program and catching a typographical error or something similar. It is usually easier, however, to trace through the program and see when it makes a computational error. At that point, one knows where the source of difficulty is and can hope to remedy it. If the basic algorithm were simple enough, it might be possible to guess the source of error by noticing a pattern in the series of mistakes it produced. Thus one might be able to find the bugs in a program--without even having a listing of it. For example,

see if you can discover the bug in the following addition program from the five sample problems.

41	328	989	66	216
+9	+917	+52	+887	+13
50	1345	1141	1053	229

Of course, if you don't have a listing of the program, you can never be certain that you have the right bug. However, you can substantiate your guess by predicting in advance the mistakes that the program would make on other problems. For example, if you have identified the bug which resulted in the answers in the previous five problems, you might want to predict the answers to the following two:

446	201
+815	+399

This particular bug is rather straightforward. We can get the same answers as the program for each of the five sample problems by "forgetting" to reset the "carry register" to zero: after doing an addition which creates a carry in a column, simply add the carry to each column to the left of it. For example, in the second problem, 8 + 7 = 15, so we get a carry of 1 into the second column. That gives us a sum of 4. If the 1 is still carried to the third column, that gives us 1 + 3 + 9 = 13. The same difficulties arise all the way across the board. Using this bug, one would predict answers of 1361 and 700 to the two extra problems.

Now the point is that a student might have this "bug" in his own arithmetic procedure, just as the computer program might. In fact, a child might well use his fingers to remember the carry, and simply forget to bend the fingers back after each carry is added. This would produce exactly the bug above.

The above is far more than an exercise in cleverness: it has tremendous implications for the way we teach. The naive view of teaching is that the teacher's obligation is to present the correct procedure coherently and well, and that if anything goes wrong, it is simply because the students have not yet succeeded in learning that procedure. The above example (and many more in the text) suggest that something very different is happening. Suppose a student is making consistent mistakes. The teacher who can diagnose such a bug in that student stands a decent chance of being able to remedy it. The teacher who looks at the student's mistakes and concludes from them simply that the student has not yet learned the correct procedure, is condemned simply to repeat the correct procedure--with much less likelihood that the student will perceive his own mistake and begin to use the correct procedure as he is supposed to.

If one makes the assumption that student's behavior is consistent when it is wrong, then the issue appears to be theoretically simple. You begin with the correct procedure, and then at each step generate what might be considered plausible bugs. Then you create a series of test problems so that the student's answers to those problems indicate which bugs he has. Having identified the bugs, you intervene directly to remedy them.

While the theory has just been made to sound remarkably simple, the implementation is actually quite complex. First, it is a surprisingly complicated task to write down all the operations that one has to do to add or subtract two three digit numbers. Primitive operations involved in subtraction, for example, include knowing the difference between any two single digits, being able to compare two digits, knowing when it is appropriate to borrow, being able to borrow, knowing to perform operations on the columns in sequence

from right to left, and many, many more. Any flaw in any one of these procedures causes a bug which needs to be diagnosed; flaws in more than one procedure cause compound bugs which may be even more difficult to diagnose. For example, the following table lists nine common procedural mistakes in the simple subtraction algorithm. When one considers possible combinations of these, things start to get out of hand very rapidly.

143 -28 125	The student subtracts the smaller digit in each column from the larger digit regardless of which is on top.
143 -28 125	When the student needs to borrow, he adds 10 to the top digit of the current column without subtracting 1 from the next column to the left.
1300 -522 878	When borrowing from a column whose top digit is 0, the student writes 9 but does not continue borrowing from the column to the left of the 0.
140 -21 121	Whenever the top digit in a column is 0, the student writes the bottom digit in the answer; i.e., 0-N = N.
140 -21 120	Whenever the top digit in a column is 0, the student writes 0 in the answer; i.e., 0-N = 0.
1300 -522 788	When borrowing from a column where the top digit is 0, the student borrows from the next column to the left correctly but writes 10 instead of 9 in this column.
321 -89 231	When borrowing into a column whose top digit is 1, the student gets 10 instead of 11.
662 -357 205	Once the student needs to borrow from a column, s/he continues to borrow from every column whether s/he needs to or not.
662 -357 115	The student always subtracts all borrows from the leftmost digit in the top number.

FIG. 1. Manifestations of some subtraction bugs.

The preceding discussion was based on the premise that students do indeed follow certain consistent procedures. Brown and Burton were able to

test this theory empirically when they were given the scores of 1325 students on a 15-item subtraction test. (By this time, they had completely automated the process for analyzing bugs, and it was done on a computer.) Their data indicates that more than 40 percent of the errors made on the test were clearly attributable to "buggy" behavior. In particular, more than 20 percent of the solution sheets they had were entirely consistent with one of the bugs they had identified. (That is, all of the answers were exactly what that particular faulty algorithm would produce.) Another 20 percent of the solution sheets indicated behavior which was strongly consistent but not identical with such a bug. This is incredibly strong evidence in favor of their thesis.

Further, the analysis of the student's performance on this test provided an interesting commentary on the test itself (which Burton and Brown had not designed). It turned out that 107 of the 1325 students tested had a bug in their "borrow from zero" procedure. In consequence, they had missed all 6 of the 15 problems on the test which called for borrowing from zero. In the original interpretation of the data, those 107 students were simply identified as students who scored 60 percent. In this analysis, they are identified as students who have not yet mastered the technique of borrowing from zero. There is quite a difference.

While the authors have computerized the analysis technique, their intent is not to have a "computer-analyst" in every classroom. Debugging is a very subtle process, and the child who is having arithmetic difficulties is very unlikely to be able to understand the explanations given him by a computer! Rather, the authors want to sensitize teachers to the idea that such consistent mistakes do occur in their students, so that the teachers themselves can do the debugging in the classroom. To this end they have written a program called

"BUGGY" which simulates a student who makes consistent arithmetic errors. Teachers can "converse" with BUGGY, and try to figure out what BUGGY is doing wrong. The following is a transcript of a dialogue between BUGGY and a team of teachers. The dialogue includes some commentary about important issues. Lines which were typed by the team of teachers are marked with a vertical line in the left margin.

Transcript of a Team Using BUGGY

WELCOME TO BUGGY.

I HAVE CHOSEN A BUG. HERE IS AN EXAMPLE OF THE BUG.

```
 17
+5
---
 13
```

The system presents an example in which the student does not get the right answer.

NOW YOU CAN GIVE ME PROBLEMS TO DETERMINE THE BUG.

```
 18      43
+6     +79
---    ---
 15     23
```

The team generates problems that are answered by the "student." Notice that the team's test problems are similar to the initial example generated by BUGGY. This is usually a poor diagnostic strategy as it may leave other symptoms of a bug undiscovered.

Got the bug!

The team forms an initial hypothesis and tells the system they have discovered the bug. The system now asks the team to describe the bug and then tests them on their understanding. Asking for a description forces the team to verbalize their sometimes fuzzy idea of what the bug is and often makes them realize how ill-defined or vague their idea actually is. In the present system, the descriptions are saved but no attempt is made by the system to understand them.

PLEASE DESCRIBE THE BUG.

Student adds the number to be carried into the last column worked in.

This cryptic description is typical of those we have observed.

Describing bugs in English is surprisingly difficult. What the team is trying to say is that if a column adds to more than 9, the units digit and the carry digit are added together and the result is written as the answer. For example, in the first column of the example problem, when 7 and 5 are added, 1 (the carry) and 2 (the units digit) are added to get 3. The same with 8 + 6, 2 + 9, 3 + 9. (This bug could easily arise if the student were using Method 2 from Fig. 1 and mistakenly wrote the carry under the same column.)

The hypothesis does account for the observed behavior. Unfortunately, it is not the student's bug. Notice the possibly humorous, probably disastrous results that would have transpired if the team had stopped at this point and tried to tutor the student on the basis of their initial belief!

HERE ARE SOME PROBLEMS TO TEST YOUR THEORY ABOUT THE BUG.

WHAT IS:

```
  21
 +39
 ---
  51
```

To ensure that the team has found the correct bug, the system produces several problems for the team to solve.

The team's hypothesis is revealed as incorrect by the first of the five test problems. BUGGY presents its answer and then allows the team to enter more problems.

THAT'S NOT THE BUG I HAVE.
USING MY BUG:

```
  21
 +39
 ---
  15
```

TRY GIVING ME SOME MORE EXAMPLES.

```
   51     99     68
 +1707  +99    +9
 ----   ---    ---
   21     36     23
```

After failing the test, the team gives a greater range of problems, specifically those using bigger numbers. How is it that 51 + 1707 can add up to 21? By broadening the scope of example problems they are able to reevaluate their initial hypothesis, dismiss it, and develop another one that is complete and consistent with respect to the new errors.

Got the Bug!

PLEASE DESCRIBE THE BUG YOU THINK I HAVE.

The student is adding all the digits together.

At this point the computer generates five more problems which the team answers correctly. The machine then congratulates the team, gives its description of the bug, and goes on to another.

There is more to BUGGY, but I suggest you go to Brown and Burton for additional details.

3. A Look at "Simple" Word Problems.

For a number of years, a group at the University of Massachusetts at Amherst has been studying a variety of students' misconceptions in college-level physics and mathematics. This discussion is based primarily on two of their working papers, "Translating Between Symbol Systems: Isolating a Common Difficulty in Solving Algebra Word Problems" by John Clement, Jack Lochhead and Elliot Soloway, and "Solving Algebra Word Problems: Analysis of a Clinical Interview" by John Clement. These papers deal with college-level students, and (at least at first) with subject matter "appropriate" for students at this level. Yet, there are two very strong similarities between this work and the work described in section 1. First, we will see again that a process which is "simple" to do correctly may yet be a rich source of potential errors. Second, there will once again be an almost remarkably perverse consistency in the way that students make mistakes--to the point where remediation is rather difficult, even if one understands what the student is doing. Finally, there will be an interesting contrast between the "static" nature of mathematical language and the "dynamic" nature of a programming language.

Since they were dealing with college-level students, the authors began with problems of some complexity. One problem, for example, asked the student to determine what price, P, to charge adults who ride a ferry boat, in order to have an income on a trip of D dollars. The students were given the following information: There were a total of L people (adults and children) on the ferry, with 1 child for each 2 adults; children's tickets are half price. The students were asked to write their equation for P in terms of the variables D and L. When fewer than 5 percent of the students given the problem solved it correctly, the authors began to use simpler and simpler problems. After a sequence of increasingly easier problems, they wound up using problems like the ones given in Table 1.

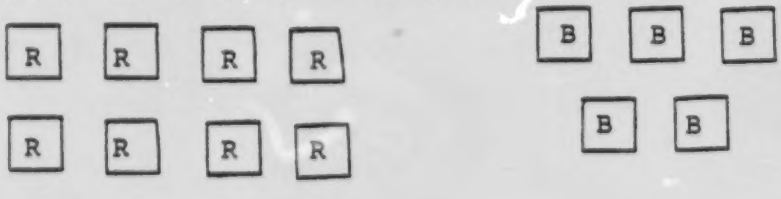
Table 1.

1. Write an equation using the variables S and P to represent the following statement: "There are six times as many students as professors at this University." Use S for the number of students and P for the number of professors.

2. Write an equation using the variables C and S to represent the following statement: "At Mindy's restaurant, for every four people who ordered cheesecake, there are five people who ordered strudel." Let C represent the number of cheesecakes and S represent the number of strudels ordered.

3. Write a sentence in English that gives the same information as the following equation: A = 7S. A is the number of assemblers in a factory. S is the number of solderers in a factory.

4. Spies fly over the Norun Airplane Manufacturers and return with an aerial photograph of the new planes in the yard.



They are fairly certain that they have photographed a fair sample of one week's production. Write an equation using the letters R and B that describes the relationship between the number of red airplanes and the number of blue planes produced. The equation should allow you to calculate the number of blue planes produced in a month if you know the number of red planes produced in a month.

The correct answers for these four problems are, of course, the following: (1) S = 6P, (2) 5C = 4S, (3) "There are 7 assemblers for every solderer," (4) 5R = 8B. The success rates for these four problems were respectively 63, 27, 29, and 32 percent.

It might seem at first that the researchers had simply found a bunch of students who were extremely defective in their algebraic skills. However, the students had been given the six questions given below in Table 2.

Table 2.

1. Solve for x: $5x = 50$
2. Solve for x: $\frac{6}{4} = \frac{30}{x}$
3. Solve for x in terms of a: $9a = 10x$
4. There are 8 times as many men as women at a particular school. 50 women go to the school. How many men go to the school?
5. Jones sometimes goes to visit his friend Lubhoft driving 60 miles and using 3 gallons of gas. When he visits his friend Schwartz, he drives 90 miles and used _?_ gallons of gas. (Assume the same driving conditions in both cases.)
6. At a Red Sox game there are 3 hotdog sellers for every 2 Coke sellers. There are 40 Coke sellers in all. How many hotdog sellers are there at this game?

On average, more than 95 percent of these problems were solved correctly. Therefore, their difficulties were not in simple algebraic manipulations. They were, rather, in translating a statement from a sentence into a suitable algebraic form.

Actually, the situation is even worse than this, for the students were more than competent in algebra. Clement's paper provides a detailed analysis of the transcript of a problem-solving session with one student who consistently gets problems of this nature wrong. The student was doing B+ work in a standard calculus course at the time of the interview, and had been able to differentiate the function $f(x) = \sqrt{x^2 + 1}$ rapidly, using the chain rule, without difficulty.

We shall examine "students and professors" problem from Table 1 in some detail. As in the Brown and Burton work, the key to the analysis is the fact that the students' errors were remarkably consistent for all of the problems in Table 1. More than four-fifths of the incorrect solutions to problem 1 were of the form 6S = P; of problem 2, of the form 4C = 5C; of problem 3, of the form "Seven solderers for every assembler"; and of problem 4, 8R = 5B. In other words, there was a very consistent reversal of the symbols and their role in the equations.

The authors identified two major causes for the reversal. The first is what we might call a "syntactic" translation of a sentence into algebraic form. The student reads along the sentence, replacing words where appropriate by algebraic symbols. Thus, "six times as many students" becomes 6S; "as" becomes equals, and "professors" becomes P. The resulting equation is 6S = P. Clinical interviews substantiated the fact that many students solved the problem this way.

The second group of students recognized that an equation does stand for a relationship between two quantities. However, the way that they represented that relationship to themselves resulted in a reversal. Many of the students, for example, drew pictures such as the one given in Figure 2 below. On one side of the desk is the professor; on the other side are the 6 students. This

is the equality: 6S = P.

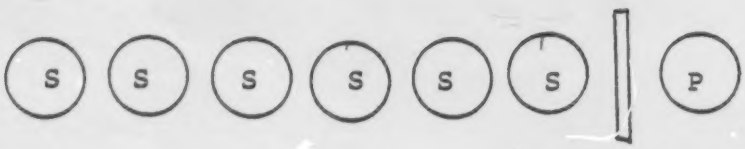


FIG. 2. Student's sketch.

To the mathematician, an equation for the "students and professors" problem is a device which allows him to calculate the number of students given the number of professors, or vice-versa. Since there are 6 times as many students as professors, one must multiply the number of professors by 6 to get the number of students (for example, 10 professors yield 60 students). Thus, S = 6P. Obviously, students do not have this perspective.

In one last experiment, the authors provide some dramatic evidence of the difference between the static and dynamic interpretations of an equation. Their "subjects" were 17 professional engineers who had between 10 and 30 years of experience each. The engineers had come to take a course in the BASIC programming language. On the first day of the course, the engineers were asked to write an equation for the following statement:

> "At the last football game, for every four people who bought sandwiches, there were five who bought hamburgers."

Only 9 out of 17 of the engineers solved the problem correctly. The following day, without there having been any discussion of the previous problem and the solution to it, the engineers were asked to write a computer program as follows:

> "At the last company cocktail party, for every 6 people who drank hard liquor, there were 11 people who drank beer. Write a program in BASIC which will output the number of beer drinkers when supplied with the number of hard liquor drinkers."

All 17 of the engineers solved the problem correctly.

The authors further substantiated these results(with less impressive, though still "significant" statistics) with a study of some college students in a programming course. Again, these results are quite surprising: it would appear a harder task to write a program (involving syntax statements, etc.) than it would be simply to write an equation. The notion of programming also suggests a possible means of remediation: if we train students to think of an equation as a "program" with inputs and outputs, we may increase the likelihood of their getting the correct answers.

4. A Look at Problem Solving.

In the preceding three sections, we saw that apparently random problem-solving behavior can actually be quite consistent. In the work with BUGGY and with elementary word problems, the focus was on consistent patterns of mistakes, for purposes of diagnosis and remediation. In this section we look at the flip side of the coin. Just as a look beneath the surface discloses consistency in novices' incorrect behavior, a look beneath the surface will also disclose great consistency in the problem-solving behavior of experts. This idea is not new, of course. It is the keystone of two major (and often irreconcilable) approaches to problem solving in this century. First, we have Polya's work on heuristics. Polya describes many strategies which, in spite of idiosyncratic differences in personal behavior, are common to expert mathematicians when they work on problems. On the opposite side of the fence, we have work in artificial intelligence. Here, too, the assumption is that problem-solving behavior can be so consistent that it can actually be modeled in enough detail for computer implementation.

Of course there are gross differences between the two approaches. Polya's work is sometimes vague, generally descriptive, and covers the entire breadth of

mathematical problem solving. Workers in artificial intelligence reject the vagueness of Polya's work, and talk about the precision in their own work: after all, they write programs which actually solve problems, and "the proof is in the pudding." However, the price they pay is that (until now at least), to obtain the precision they need, they must work in extremely narrow subject area domains.

My work is an attempt to reconcile these two approaches, with what might be called a type of "human artificial intelligence." As in artificial intelligence, we might try to model expert behavior. However, the goal is not to model it for machine implementation. The idea instead is to pick out those aspects of expert behavior which students can learn. The discussion which follows is a distilled version of my paper "Teaching Problem Solving Skills," which will appear later this year in the Monthly. Other papers which deal with the same ideas in more detail are given in the bibliography.

To make the point that experts and novices approach problems in dramatically different ways, consider the following three problems--all of which are ostensibly accessible to high school students.

Problem 1: *Let a, b, c, and d be given numbers between 0 and 1. Prove that* $(1-a)(1-b)(1-c)(1-d) > 1-a-b-c-d$.

Problem 2: *Determine the sum* $\frac{1}{2!} + \frac{2}{3!} \cdots + \frac{n}{(n+1)!}$.

Problem 3: *Prove that if* $2^n - 1$ *is a prime, then n is a prime.*

On problem 1 most students will laboriously multiply the four factors on the left, subtract the terms on the right, and then try to prove that (ab+ac+ad+bc+bd+cd-abc-abd-acd-bcd+abcd) > 0--usually without success. Virtually all the mathematicians I've watched solving it begin by proving the

inequality (1-a)(1-b) > 1-a-b. Then they multiply this inequality in turn by (1-c) and (1-d) to prove the three- and four-variable versions of it.

Likewise in problem 2, most students begin by doing the addition and placing all the terms over a common denominator. A typical expert on the other hand, begins with the observation "That looks messy. Let me calculate a few cases." The inductive pattern is clear and easy to prove.

The colleague who read problem 3 and said "That's got to be done by contradiction" was typical; given the structure of the problem, one really has no alternative. Yet this almost automatic expert observation is alien to students: a large number of those to whom I have given the problems either respond with comments like "I have no idea where to begin" or try a few calculations to see whether the result is plausible, then reach a dead end.

Of course these are special problems, for which expert and novice performance is remarkably consistent. While the experts did not consciously follow any strategies, their behavior was at least consistent with these "heuristic" suggestions:

a. For complex problems with many variables, consider solving an analogous problem with fewer variables. Then try to exploit either the method or the result of that solution.

b. Given a problem with an integer parameter n, calculate special cases for small n and look for a pattern.

c. Consider argument by contradiction, especially when extra "artillery" for solving the problem is gained by negating the desired conclusion.

Many of the novices were unaware of the strategies, and many others "knew of them" (that is, upon seeing the solution they acknowledged having seen similar

solutions] but hadn't thought to use them. Expert and novice problem solving are clearly different. The critical question is: Can we train novices to solve the problems as experts do?

There are a number of obstacles. First, we have to factor out simple subject matter knowledge: there is no way that one can hope to give the students experience before they have it, or to compensate for it. Rather, we would like to provide the students with strategies for approaching problems with flexibility, resourcefulness, and efficiency.

Second, we must realize that the heuristic strategies described by Polya are far more complex than their descriptions would at first have us believe. Consider the following strategy and a few problems.

"To solve a complicated problem, it often helps to examine and solve a simpler analogous problem. Then exploit your solution."

Problem 4: Two points on the surface of the unit sphere (in 3-space) are connected by an arc A which passes through the interior of the sphere. Prove that if the length of A is less than 2, then there is a hemisphere H which does not intersect A.

Problem 5: Let a, b, and c be positive real numbers. Show that not all three of the terms a(1-b), b(1-c), and c(1-a) can exceed 1/4.

Problem 6: Find the volume of the unit sphere in 4-space.

Problem 7: Prove that if $a^2 + b^2 + c^2 + d^2 = ab+bc+cd+da$, *then* $a=b=c=d$.

These four problems, like problem 1, can be solved by the "analogous problem" strategy. Yet, it is unlikely that a student untrained in using the

strategy would be able to apply it successfully to many of these. Part of the reason is that the strategy needs to be used differently in the solution of each problem.

In solving problem 1, we built up an inductive solution from the two-variable case, using the result of the analogous problem as a stepping stone in the solution of the original.

In contrast, analogy is used in problem 4 to furnish the idea for an argument. The problem is hard to visualize in 3-space but easy to see in the plane: we want to construct a diameter of a unit circle which does not intersect an arc of length 2 whose endpoints are on the circle. Observing that the diameter parallel to the straight line between the endpoints has this property enables us to return to 3-space and to construct the analogous plane.

Problem 5 is curious. It looks as though the two-variable analogy should be useful, but I haven't found an easy way to solve it. At first the one-variable version looks irrelevant, but it's not. If you solve it, and think to take the product of the three given terms, you can solve the given problem. So again we exploit a result, but this time a different result in a different way.

Problem 6 exploits both the methods and results of the lower-dimensional problems. We integrate cross-sections, using the same method; the measures of the cross-sections are the results we exploit.

In problem 7 it would seem apparent that the two-variable problem is the appropriate one to consider. However, which two-variable problem is not at all clear to students. A large number of those I have watched tried to solve

Problem 7': Prove that $a^2 + b^2 = ab$ *implies that* $a=b$, *instead of*
Problem 7": Prove that $a^2 + b^2 = ab + ba$ *implies* $a = b$.

We conclude that the description "exploiting simpler analogous problems" is really a convenient label for a collection of similar, but not identical, strategies. To solve a problem using this strategy, one must (a) think to use the strategy (this is non-trivial!), (b) be able to generate analogous problems which are appropriate to look at, (c) select among the analogies for the appropriate one, (d) solve the analogous problem, and (e) be able to exploit either the method or result of the analogous problem appropriately.

If we assume now that we can actually describe the strategies in enough detail so that people can use them, we run right into another problem. That is: a list of all the strategies in detail would be so long that the students could never use it! Knowing how to use the strategy isn't enough: the student must think to use it when it is appropriate.

Consider techniques of integration in elementary calculus. There are fewer than a dozen important techniques, all of them algorithmic and relatively easy to learn. Most students can learn integration by parts, or substitution, or partial fractions, as individual techniques and use them reasonably well, as long as they know which techniques they are supposed to use. (Imagine a test on which the appropriate technique is suggested for each problem. The students would probably do very well.) When they have to select their own techniques, however, things often go awry. For example, $\int \frac{x\,dx}{x^2-9}$, a "gift" first problem on a test, caused numerous students trouble when they tried to solve it by partial fractions or, even worse, by a trigonometric substitution!

In "Presenting a Strategy for Indefinite Integration" (The Monthly, Oct. 1978) I discuss an experiment in which half the students in a calculus class (not mine) were given a strategy for selecting techniques of integration, based on a model of "expert" performance. The other students were told

to study as usual--using the miscellaneous exercises in the text to develop their own approaches to problem solving. Average study time for members of the "strategy" group was 7.1 hours, while for the others it was 8.8 hours; yet the "strategy" group significantly outperformed the rest on a test of integration skills--in spite of the fact that they were not given training in integration, just in selecting the techniques of integration.

The "moral" to the experiment is that students who cannot choose the "right" approach to a problem--even in an area where there are only a few useful straightforward techniques--do not perform nearly as well as they "should." If we leap from techniques of integration to general mathematical problem solving, the number of potentially useful techniques increases substantially, as does the difficulty and subtlety in applying the techniques. An efficient means for selecting approaches to problems, for avoiding "blind alleys," and for allocating problem-solving resources in general thus becomes much more critical. Without it, the benefits of training in individual heuristics may be lost.

In consequence of the above, an attempt to teach general mathematical problem solving would need these two components: first, a detailed description of individual strategies, and second, a global framework for selecting these strategies and using them efficiently. One way of presenting such a framework is with a "model" of expert problem solving. That model takes a semester to unfold, so there is no sense in my attempting to summarize it here. What I have done is simply to give the outline of the model (see Figure 3), and a description of the most important heuristic strategies which fall within each of the major blocks of that strategy (see Figure 4).

SCHEMATIC OUTLINE OF THE PROBLEM-SOLVING STRATEGY

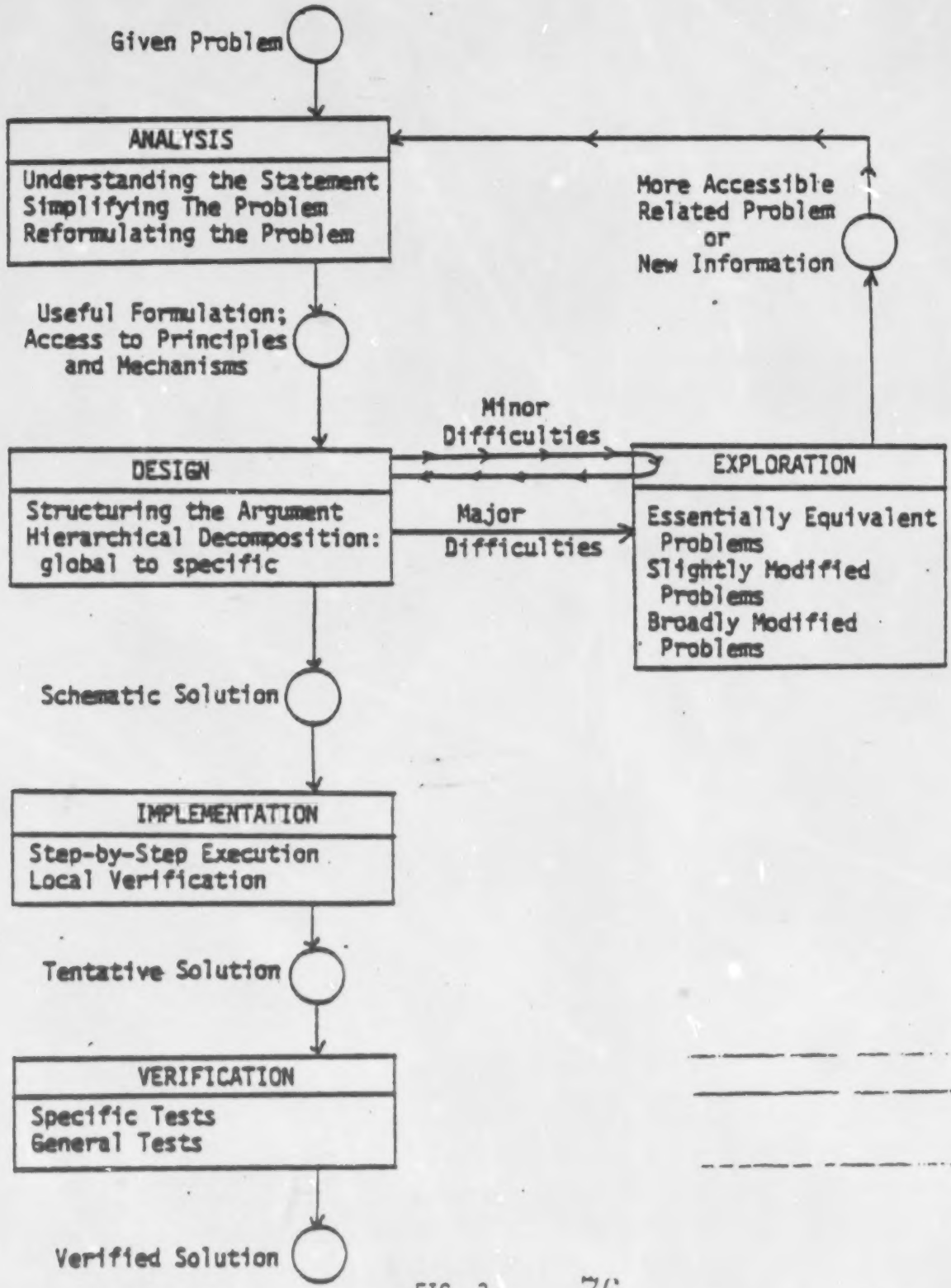


FIG. 3

SOME IMPORTANT HEURISTICS IN PROBLEM SOLVING

For Analyzing and Understanding A Problem:

1. Draw a Diagram if at all possible.
2. Examine Special Cases, (a) to exemplify the problem, (b) to explore the range of possibilities through limiting cases,(c) to find inductive patterns by setting integer parameters equal to 1,2,3,... in sequence.
3. Try to simplify it, by using symmetry or "without loss of generality."

For the Design and Planning of a Solution:

1. Plan solutions hierarchically.
2. Be able to explain, at any point in a solution, what you are doing and why; what you will do with the result of this operation.

For Exploring Solutions to Difficult Problems:

1. Consider a variety of equivalent problems,
   (a) replacing conditions by equivalent ones,
   (b) recombining elements of the problem in different ways,
   (c) Introducing auxiliary elements,
   (d) Re-formulating the problem by (i) a change of perspective or notation, (ii) arguing by contradiction or contrapositive, or (iii) assuming a solution and determining properties it must have.
2. Consider Slight modifications of the original problem:
   (a) Choose subgoals and try to attain them.
   (b) Relax a condition and try to re-impose it.
   (c) Decompose the problem and work on it case by case.
3. Consider Broad modifications of the original Problem:
   (a) Examine analogous problems with less complexity (fewer variables).
   (b) Explore the role of just one variable or condition, the rest fixed.
   (c) Exploit any problem with a similar form, "givens," or conclusions; try to exploit both the result and the method.

For Verifying a Solution:

1. Use these specific tests: does it use all the data, conform to reasonable estimates, stand up to tests of symmetry,dimension analysis,scaling?
2. Use these general tests: Can it be obtained differently, substantiated by special cases, reduced to known results, generate something you know?

FIG. 4

Of course, documenting improved problem-solving ability is rather difficult. I am slowly amassing evidence, in a variety of different ways, that instruction in problem solving actually can have an impact on students' problem-solving performance. The material on integration provided some evidence of this. A "laboratory study" demonstrated that "problem-solving experience" in and of itself is not enough: in the experiment, two groups of students worked on the same problems for the same amount of time and saw the same solutions, but one saw in addition heuristic explanations of the solutions. The differences in their performances were dramatic. (See "Explicit Heuristic Training as a Variable in Problem Solving Performance.") Third, there is a large amount of "before and after" data on the students in the problem-solving course. These data indicate both an improved problem-solving performance on the part of the students and an improved ability to generate plausible approaches to problems, as opposed to a control group. (These data are shaky, and I do not want to base any claims on them for fear of being lumped with the people I condemned for the inappropriate use of statistics at the beginning of this paper. Next year, I will be in a better position: I plan to teach a course virtually identical in subject matter to my problem-solving course, save for the fact that I do not discuss the problem-solving strategy in particular. That will be as close a controlled group as I can hope to get; at that point, I will have more faith in the statements I can make.) In the meantime, there is much data to be analyzed by a variety of different means--means which were unavailable just a few years ago, and which come from a variety of disparate sources. As one such example, let me discuss briefly the notion of "hierarchical cluster analysis." Consider the following three problems.

Problem 8: Given that lines intersect if and only if they are not parallel, and that any two points in the plan determine a unique line between them, prove that any two distinct nonparallel lines must intersect in a unique point.

Problem 9: Given 22 points on the plane, no three of which lie on the same straight line, how many straight lines can be drawn, each of which passes through two of those points?

Problem 10: If a function has an inverse, prove that it has only one inverse.

Let us take an extreme case. The student who understands virtually nothing of these problems may think that problems 8 and 9 are related because they both deal with lines in the plane. On the other hand, the mathematician sees that both problems 8 and 10 deal with the uniqueness, and are likely to be proved by contradiction. Therefore he may perceive of those problems as being similar.

Suppose 100 students were given these 3 problems, and asked to group together those problems which they thought were related. (They might decide that none of the problems was related or that two of them were, or that three of them were.) One could then create a 3 by 3 matrix, where the i,j-th entry represented the number of students who considered the i-th and j-th problems to be related. A comparison of these matrices before and after instruction, for both experimental and controlled groups, could indicate changes in the students' perceptions of the way these problems were structured mathematically.

In fact, my cluster analysis used 32 problems, with a 32 x 32 matrix for analysis. There were clear differences between experimental pre- and post-test scores, and controlled pre- and post-test scores. Further comparison with "expert" sorting of the problems is also planned. The full tally is yet to come, but the preliminary results are encouraging. There will be more about techniques such as this in the next section.

5. Summary and Conclusion.

The three pieces of work I described above give barely a taste of research in mathematics education today. It is interesting to note that none of the people involved comes from the "classical" mathematics education community. Brown and Burton come from what might be loosely described as the "artificial intelligence" community. Clement, Lochhead and Soloway are housed in the Physics Department at Amherst. I came from pure mathematics, and the techniques I use are derived from both AI and modern cognitive psychology. My work has profited greatly from criticism by classical math educators and members of all the communities mentioned above. This is indicative, I think, of the general state of education today: research now is highly interdisciplinary, and profits greatly from being so.

The one problem with such breadth and scope is that results are scattered and hard to find. I found out about all the work above, for example, from the research "grapevine" in mathematics education. Few mathematicians read Cognitive Science, for example, or the International Journal for Man-Machine Studies. Publications of the AMS for the most part do not deal with mathematics education; space in the Monthly is severely limited, and is reserved for articles directly appropriate to college-level math education; the Journal

for Research in Mathematics Education is not read by mathematicians for the most part, but by math educators; therefore it is difficult for the mathematical community to have a good idea of the new and exciting things in mathematics education. I hope the preceding has whetted your appetite. If so, you will find many more interesting things to read in the bibliography. I have not at all tried to be comprehensive; in fact, I have omitted those sources with which the mathematician is likely to be familiar. If you like what you see in the bibliography, the bibliographies in those articles will lead you to more. In any case, I hope you have been convinced of one thing: there is hope.

BIBLIOGRAPHY

Bloom, B. S. Problem Solving Processes of College Students. Chicago: University of Chicago Press, 1950.

Brown, John Seely and Burton, Richard R. Diagnostic Models for Procedural Bugs in Basic Mathematical Skills. Cognitive Science, 2: 155-192 (1978), Cambridge, Massachusetts.

_____. A Paradigmatic Example of an Artificially Intelligent Instructional System. Int. J. Man-Machine Studies (1978) 10, 323-339, Cambridge, Massachusetts.

Bundy, Alan. Analyzing Mathematical Proofs (or Reading Between the Lines). D.A.I. report #2, University of Edinburgh (1975).

Burton, Richard R. and Brown, John Seely. An Investigation of Computer Coaching for Informal Learning Activities. Int. J. Man-Machine Studies (1979) 11, 5-24, Cambridge, Massachusetts.

Chase, William G. and Chi, Michelen T. H. Cognitive Skill: Implications for Spatial Skill in Large-Scale Environments. University of Pittsburgh Learning Research and Development Center, December 15, 1979.

Clement, John. "Solving Algebra Word Problems: Analysis of a Clinical Interview." Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts, March, 1979.

Clement, John, Lochhead, Jack and Soloway, Elliot. "Translating Between Symbol Systems: Isolating a Common Difficulty in Solving Algebra Word Problems." Cognitive Development Project, Department of Physics and Astronomy, University of Massachusetts, 1979.

Duncker, K. On Problem Solving. Psychological Monographs, 58, No. 5. Washington, D.C.: American Psychological Association, 1945.

Lakatos, I. Proofs and Refutations. Cambridge: Cambridge University Press, 1976. (Revised 1977)

Lesh, Richard, Editor. Applied Problem Solving.. ERIC Clearinghouse for Science, Mathematics, and Environmental Education. Columbus, Ohio (1980).

Lochhead, Jack and Clement, John, Editors. Cognitive Process Instruction. Philadelphia: Franklin Institute Press, 1980.

EPISODES AND EXECUTIVE DECISIONS IN MATHEMATICAL PROBLEM SOLVING*

Alan H. Schoenfeld
Mathematics Department
Hamilton College
Clinton, N.Y. 13323

SE038234

*This paper was presented at the 1981 AERA Annual Meeting, Los Angeles, Ca., April 1981

The research reported here was supported by a National Science Foundation Research in Science Education Grant, SED 79-19049.

Running Head: Episodes and Executive Decisions

To appear in Acquisition of Mathematics Concepts and Processes, R. Lesh & M. Landau, Eds. N.Y. Academic Press, In press.

Abstract

The research described here seeks to characterize the "managerial" aspects of expert and novice problem-solving behavior, and to describe the impact of managerial or "executive" actions on success or failure in problem solving. We present a framework for analyzing protocols of problem-solving sessions based on "episodes" of problem-solving behavior and focusing on managerial decisions between episodes. Experts are shown to have rather "vigilant" managers, which strive for efficiency and accuracy. In contrast, novices squander their problem-solving resources because they lack such managers.

Episodes and Executive Decisions in Mathematical Problem Solving

Introduction and Overview

This is a rather speculative paper dealing with "managerial" decisions in human problem solving. It presents a (still evolving) framework for the analysis at the macroscopic level of problem-solving protocols, focusing on "executive" behaviors. The paper is based on the following premise.

There are two qualitatively different kinds of decisions, which we shall call "tactical" and "strategic," which are necessary in broad, semantically rich domains (for example, mathematical problem solving at the college freshman level). The first, tactical decision making, has received the lion's share of attention. By tactics I mean "things to implement." Tactics include all algorithms and most heuristics, both of the Pólya type (e.g., draw a diagram whenever possible; consider special cases) and of the kinds used in Artificial Intelligence (means-ends analysis, hill-climbing). Given that one has decided to calculate the area of a particular region, the choice of whether to approach that calculation via trigonometry or analytic geometry is a tactical choice.

In contrast, "strategic" or managerial decisions are those which have a major impact on the direction a solution will take, and on the allocation of one's resources during the problem-solving process. For example: If one is given twenty minutes to work on a problem and calculating the area of a region is likely to take ten minutes, the decision to calculate the area of that region is a strategic one--regardless of the method ultimately chosen

for performing the calculation. Like a decision during wartime to open a front, this one choice may determine the success or failure of the entire enterprise.

This separation of managerial decisions from implementation decisions has implications for both human and machine problem solving. Mathematics problem-solving instruction to date has focused largely, and with somewhat questionable success, on heuristics or "tactics." I propose that much of the reason for this lack of success lies in the fact that attention to managerial behaviors has mostly been neglected. The protocols discussed below will indicate that heuristic fluency is of little value if the heuristics are not "managed" properly. I believe that much greater attention will have to be paid to "metaheuristics" or managerial actions in classroom instruction, if we are to be successful in teaching problem-solving skills.

There appear to be parallels in artificial intelligence. Regardless of their sophistication, production systems are essentially tactical decision makers. They are not strategists. The managerial decisions made in such programs, by "conflict resolution strategies" when the conditions for more than one production are met simultaneously, seem to be more or less ad hoc and idiosyncratic, rather than theory-based. For the most part, programming in narrow domains finesses the question of managerial strategies. However, such concerns cannot be ignored as the domains of investigation are broadened. Further, some attempt at dealing with executive strategies must be made for the creation of "glass box" experts in computer-based tutorial systems for non-trivial domains. Since such decisions are an important component of human problem solving, any system in a broad arena which ignores them will lack psychological validity.

This paper discusses a framework for examining, at the macroscopic level, a broad spectrum of problem-solving protocols. Protocols are parsed into

major "episodes." These are periods of time during which the problem solver(s) is engaged on a single set of like actions, such as "planning" or "exploration." It is precisely between such episodes that the managerial decisions which can "make or break" a solution are often made, or not made. We focus on decision making at these points, and on the impact of such decisions--or their absence--on problem-solving performance. The quality and success of problem-solving endeavors will be seen to correspond closely (in human problem solving) to the presence, and vigilance, of such "managers."

A Discussion of Antecedents

By definition, protocol coding schemes are concerned with producing objective records or "traces" of a sequence of overt actions taken by individuals in the process of solving problems. In mathematics education, the coded protocol is generally subjected to a qualitative analysis; often correlations will be sought between certain types of behavior (e.g., the presence of goal-oriented heuristics) and problem-solving success. In artificial intelligence, the goal is often to write a program that will simulate a given protocol, or the idealized behavior culled from a variety of protocols. In both cases the level of analysis is microscopic. My goal here is to indicate that in many cases the microscopic level analysis may be entirely inappropriate. In analyzing human problem solving, attention to that level of detail may cause one to "miss the forest for the trees"; if the wrong strategic decisions are made, tactical ones are virtually irrelevant. In artificial intelligence, great progress has been made at the tactical level through the use of production systems. It is not at all clear, however, that they will serve well for making managerial decisions. I believe that we may wish to think of these executive decisions as being at a higher level than tactical ones, and may want to deal with these "strategists" separately.

(Note: what follows is an opinionated discussion of the recent literature, which depends heavily on the distinction between "tactical" or "strategic" or "managerial" decisions. These distinctions may be much clearer after the reader has considered the examples discussed in the next section. Thus the reader may wish to skip ahead to that section, and later consider the comments made here in the light of those examples.)

The following description, taken from Lucas et al., (1979, p. 354) is typical of the efforts of mathematics educators to deal with problem-solving protocols.

> [T]he authors came to agreement on the definitions for a set of constructs which were to represent observable, disjoint problem solving behaviors and related phenomena . . . . Each event was assigned a symbol, and the collection of events which comprised a problem-solving sequence of processes was recorded in a horizontal string of symbols corresponding to the chronological order of appearance during the actual problem solution. In this manner a researcher could listen to a tape of a problem solution (in conjunction with observing written work, interviewer notes, and/or a verbatim transcript) and produce a string of symbols which represented the composite perception of the solution process. Conversely, an examination of the given string of symbols could be used to provide a reasonably clear picture of what had happened during a problem-solving episode.

That particular coding scheme included a two-page "dictionary" of processes which were assigned coding symbols. All behavior was "required to be explicit; otherwise it is not coded." (p. 359) As an example of the coding, the sequence (p. 361)

> The problem solver reads the problem, hesitates, rereads part of the problem, says the problem resembles another problem and he will try to use the same method, then deduces correctly

a piece of information from one of the given data was coded as $(R,R,L_8P_iD_{a5}4)$.

In part because of the cumbersome nature of such systems and the wealth of symbols that must be dealt with, once coded, other researchers have opted to focus on more restricted subsets of behaviors. Kulm's recent NSF-supported work, "Analysis and Synthesis of Mathematical Problem Solving Processes," uses a revised and more condensed process code dictionary (private communication, 1979). Kantowski's recent work (Note 3) includes a "coding scheme for heuristic processes of interest" which focuses on five heuristic processes related to planning, four related to memory for similar problems, and seven related to looking back. The frequency of such processes is related to problem-solving performance.

So far as I know, there are no systems for protocol analysis that focus in any substantive way on strategic decisions. There are no frameworks for dealing with things which ought to have been considered, but were not. For the most part, discussions in the literature of executive decision making during problem solving are weak. Polya, for example (1965, p. 96) offers "Rules of Preference" for choosing among options in a problem-solving task. These include injunctions such as "the less difficult precedes the more difficult" and "Formerly solved problems having the same kind of unknown as the present problem precede other formerly solved problems." My own attempts (Schoenfeld, 1979; 1980) at capturing a managerial strategy in flow chart form for students' implementation were somewhat impoverished, the flow chart in effect presenting a default strategy. All other factors being equal--meaning that the problem solver had exhausted the lines of attack which had appeared fruitful (his 'productions?") and had no strong leads to follow up--it was considered reasonable to try the heuristic suggestions in this "managerial strategy," roughly in

the manner suggested by the flow chart. This bypassed the tough questions, however. Issues like: how does one decide what to pursue; for how long; how does one evaluate progress towards a solution; when should the "manager" interfere, etc.,while discussed in class, were not formally a part of the strategy. Moreover, there was no systematic and rigorous framework for examining these questions.

As a result of (1) the narrowness of the problem domains in which artificial intelligence has successfully operated, and (2) the tactical utility of production systems in those domains, the AI community has given even less attention to executive strategies than has the math-ed community. The questions are not new: the "considerations at a position in problem space" listed by Allen Newell (1966, figure 5) are quite similar to those we will pose below. But

> "Select new operator:
>
> Has it been used before?
>
> Is it desirable: will it lead to progress?
>
> Is it feasible: will it work in the present situation if applied?"

takes on very different shades of meaning at the strategic rather than the tactical level. So far as I can tell, (and my knowledge of such is limited) recent advances in production systems allow for rather clever tactical decision making. There are computationally efficient means of keeping track of and sorting through productions for relevancy, and there are conflict resolutions systems (McDermott & Forgy, 1978) for selecting among productions when the conditions for more than one of them have been satisfied. Such structures

prohibit productions from executing more than once on the same data. This prevents the kind of endless reptitions all too common in students and forces, if necessary, the examination of all available information. Since preference is given to productions whose conditions are satisfied by elements most recently placed in working memory, there is a "natural" continuity to the sequence of operations. Other means of selection (e.g., specificity precedes generality) provide plausible means of selecting tactics in relatively narrow domains. Yet I am not sure that the level of analysis is right for general problem solving, or that such strategies would have much to say about the strategic decisions in the examples given in the next section. Similar comments apply to the "adaptive" or "self-modifying" production systems described by Anzai and Simon (1979), Neves (1978), and Neches (1979). While the learning principles they exemplify may be general, the embodiments of those principles in those papers are at the tactical level. Simon (1980) argues that "effective professional education calls for attention to both subject matter knowledge and general skills (p. 86)" and then goes on to say (p. 91) that "general skills (e.g., means-ends analysis) will be particularly important in the learning stages but will also show up implicitly in the form of the productions that are used in the skilled performance." But even this is one step removed from the heart of the matter: what underlies the form of the productions is in the mind of the programmer, not in the productions. We need a methodology for focusing on those general skills directly.

An Informal Analysis of Two Protocols

The AI literature is filled with beautiful protocols. I have never been that lucky: those generated by my students (and to some extent by my colleagues) in the process of grappling with relatively unfamiliar problems have been, on the whole, rather unaesthetic. This section considers two such protocols, each generated by a pair of students. (Following a suggestion from John Seely Brown, I have students work on problems in pairs. While the question "why did you do that?" coming from me may be terribly intimidating and is likely to alter the solution path, the question "why should we do that?" from a fellow student working on a problem is not. This type of dialogue between students often serves to make managerial decisions overt, whereas such decisions are rarely overt in single-student protocols.) An informal analysis, focusing on the importance of managerial decisions, follows.The formal analytic structure is given in the next section.

Protocols 1 and 2 are given in Appendices 1 and 2, respectively. The students were asked to work on the problem together, out loud, as a collaborative effort. They were not to go out of their way to explain things for the tape, if that interfered with their problem solving; their interactions, if truly collaborative, would provide me with the information I needed. (See Ericsson and Simon (1978; 1979) for a discussion of instructions for speak-aloud experiments.) All of the students were undergraduates at a liberal arts college. Students A and K (protocol 1) had 3 and 1 semesters of college mathematics (calculus) respectively. Students D and B (protocol 2) each had 3 semesters of college mathematics. It should be recalled that such students, by most standards, are successful problem solvers: the unsuccessful ones had long since stopped taking mathematics courses. Both protocols are of the same

problem:

> Three points are chosen in the circumference of a circle of radius R, and the triangle containing them is drawn. What choice of points results in the triangle with the largest possible area? Justify your answer as best you can.

If protocol 1 makes for confused reading, the tape it was taken from makes for even more pained viewing. I would summarize the problem-solving session as follows:

The students read and understood the problem, and then quickly conjectured that the answer was the equilateral triangle. They impetuously decided to calculate the area of the triangle, and spent the next 20 minutes doing so. These calculations of the area were occasionally punctuated by suggestions which might have salvaged the solution, but in each case the suggestions were quickly dropped and the students returned to their relentless pursuit of the worthless calculation. (Neither student could tell me, after the cassette ran out of tape, what good it would do them to know the area of the equilateral triangle.) Observe the following.

1. The single most important event in the twenty-minute problem-solving session, upon which the success or failure of the entire endeavor rested, was one which did not take place--the students did not assess the potential utility of their planned actions, calculating the area of the equilateral triangle. In consequence, the entire session was spent on a wild goose chase.
2. Inadequate consideration was given to the utility of potential alternatives which arose (and then submerged) during the problem-solving process. Any of these: the related problem of maximizing a rectangle in a circle (item

28), the potential application of the calculus (item 52) for what can indeed be considered a max-min problem; the qualitative varying of triangle shape (item 68) might have, if pursued, led to progress. Instead, the alternatives simply faded out of the picture. (See, for example items 27 to 31.)

3. Progress is never monitored or (re)assessed, so that there is no reliable means of terminating wild goose chases once they have begun. (This is to be strongly contrasted with an expert protocol, where the problem solver interrupted the implementation of an outlined solution with "this is too complicated. I know the problem shouldn't be this hard.")

Now, how does one code such a protocol? First, we should observe that matters of detail (such as whether or not the students will accurately remember the formula for the area of an equilateral triangle, items 73 to 75) are virtually irrelevant. To return to the military analogy in the opening section: if it was a major strategic mistake to open a second front in a war, the details of how a hill was taken in a minor skirmish on that front are of marginal interest.

A second and more crucial point is that the overt actions taken by the problem solvers in that protocol are, in a sense, of minor import. The problem-solving effort was a failure because of the absence of assessments and strategic decisions. Any framework that will make sense of that protocol must go beyond simply recording what did happen; it should suggest when strategic decisions ought to have been made, and allow one to interpret success or failure in the light of whether, and how well, such decisions were made.

If protocol 1 stands as evidence of the damage that can be caused by a manager "in absentia," protocol 2 provides evidence of the catastrophic effects of bad management. The processes in this tape were not muddled, as in protocol 1; the decisions were overt and clear. The next paragraph summarizes the essential occurrences in the tape. The superscripts refer to the commentary that follows.

D and B quickly conjecture that the solution is the equilateral triangle, and look for ways to show it. D, apparently wishing to exploit symmetry in some way, suggests that they examine triangles in a semicircle with one side as diameter. They find the optimum under these constraints, and reject it "by eye" as inferior to the equilateral.[1] Still focusing on symmetry, they decide[2] to maximize the area of a right triangle in a semicircle, where the right angle lies on the diameter. This (serendipitously correct) decision reduces the original problem to a 1-variable calculus problem[3] which B proceeds to work on. Twelve minutes later the attempt is abandoned,[4] and the solution process degenerates into an aimless series of explorations, most of which serve to rehash the previous work.[5]

1. Rejecting the alternative is quite reasonable, as are their actions in analyzing the problem up to this point. However, this blanket rejection may have cost them a great deal. The variational argument they used to find the isosceles right triangle (holding the base fixed and observing that the area is largest when the triangle is isosceles) is perfectly general and can be used to solve the original problem as stated. But the students simply turn away from their unsuccessful attempt, without asking if they could learn from it. In doing so, they may have "thrown out the baby with the bath water."

2. This decision, which affects the direction of the solution for more than 60% of the allotted time, is made in a remarkably casual way (items 24 to 27):

   D: (after one attempt at symmetry has failed) ...you want to make it perfectly symmetrical, but we can, if we maximize this area, just flip it over, if we assume that it is going to be symmetrical.

   B: Yea, it is symmetrical.

   This assumption is not at all justified (they are assuming part of what they are to prove). The students have changed the problem and proceed, without apparent concern, to work on the altered version.

3. B's tactical work here is quite decent, as is much of both students' tactical work throughout the solution process. The decision to "scale down" the problem to the unit circle (item 37) is just one example of their proficiency. There is awareness of, and access to, a variety of heuristics and algorithmic techniques during the solution. Unfortunately, B lost a minus sign during this particular calculation, which gave him a physically impossible answer. He was aware of it; local assessment worked well. However, global assessment (see 4 and 5) did not.

4. This decision to abandon the analytic approach is just as astonishing, in the way it takes place (items 74 and 75) as the decision to undertake it:

   D: Well, let's leave the numbers for a while and see if we can do it geometrically.

   B: Yea, you're probably right.

   Given that more than 60% of the solution has been devoted to that approach (and that correcting a minor mistake would salvage the entire operation),

this casual dismissal of their previous efforts has rather serious consequences.

5. There were a number of clever ideas in the earlier attempts made by D and B. Had there been an attempt at a careful review of those attempts, something might have been salvaged. Instead, there was simply a "once over lightly" of the previous work that added nothing to what they had already done.

A framework for focusing on the managerial decisions in such protocols is discussed in the next section.

A (poorly defined and still evolving) Framework for the Macroscopic Analysis of Certain Kinds of Problem-Solving Protocols

The two protocols discussed in the preceding section raise the major questions I wish to address here. I believe that decisions at the managerial level may "make or break" a problem-solving attempt, and that (at least in the case of poor managerial decisions) these may render irrelevant any subsequent tactical (i.e., implementation) decisions. Thus we focus on behavior at the macroscopic level.

Protocol 1, which is rather typical of students' problem solving, illustrates one of the major difficulties in dealing with managerial decisions: the absence of intelligent management may doom problem-solving attempts to failure. Yet all extant schemes focus on what is overtly present, ignoring the crucial decisions that might (and should!) have taken place. Protocol 2 is, in a sense, easier to deal with. The decisions were overt, though poor. This protocol serves to indicate that decision making means more than simply choosing solution paths: it incorporates local and global assessments of progress, as well as trying to salvage the valuable elements of ultimately flawed approaches. This section offers a scheme for parsing protocols that tries to address these issues.

There are both objective and subjective components to the framework for analyzing protocols. The objective part consists of identifying, in the protocol, the loci of potential managerial decisions. The subjective part consists of characterizing the nature of the decision-making process at these "managerial decision points" and describing the impact of those decisions (or their absence!) on the overall problem-solving process.

By definition, managerial or strategic action is appropriate whenever a large amount of tactical resources are about to be expended. This provides the basic idea for parsing the protocols. Partition a protocol into macroscopic chunks of consistent behavior ("episodes"). Then the points between episodes--where the direction or nature of the problem solution changes significantly--are the managerial decision points where, at minimum, managerial action ought to have been considered.

In addition to these junctures between episodes, there are two other loci for managerial action: at the arrival of new information or the suggestion of new tactics, and at the point where a series of tactical failures indicates that strategic review might be appropriate. The loci that deal with new information are well defined and pose little difficulty in identification. Observe that this kind of decision point can occur in the middle of an episode: new information may be ignored or dismissed (at least temporarily), and the problem solver may continue working along previously established lines. The latter kind is more difficult, and calls for subjective judgment; I have no easy way of dealing with these at present. At some point when implementation bogs down, or when the problem-solving process degenerates into more or less unstructured explorations, it is time for an "executive review." It is clear from the protocols I have taken that experts have "monitors" that call for such review, and that novices often lack them. We will return to this point later, in the subjective analysis.

Figures 1 and 2 represent a parsing of protocols 1 and 2, respectively, into episodes. "New information" points within episodes are indicated.

Insert Figures 1 and 2 about here

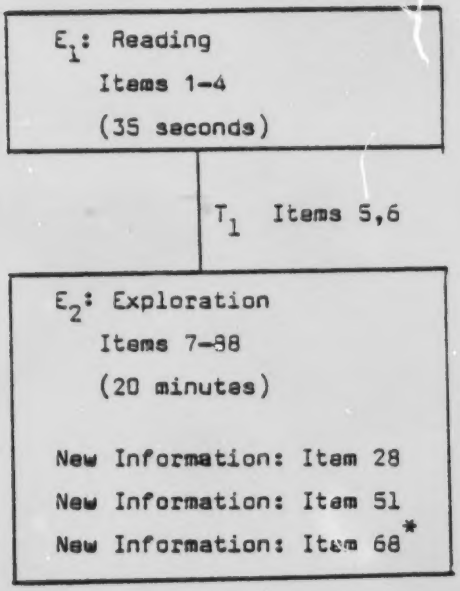


Figure 1

A Parsing of Protocol 1

*Note: From the written protocol it might appear that Item 68 begins a new episode. In fact, the students had lost virtually all their energy by that point, and were merely doodling; they returned (after the tape clicked off) to musings about the equilateral triangle. Thus items 6-88 are considered to be one episode.

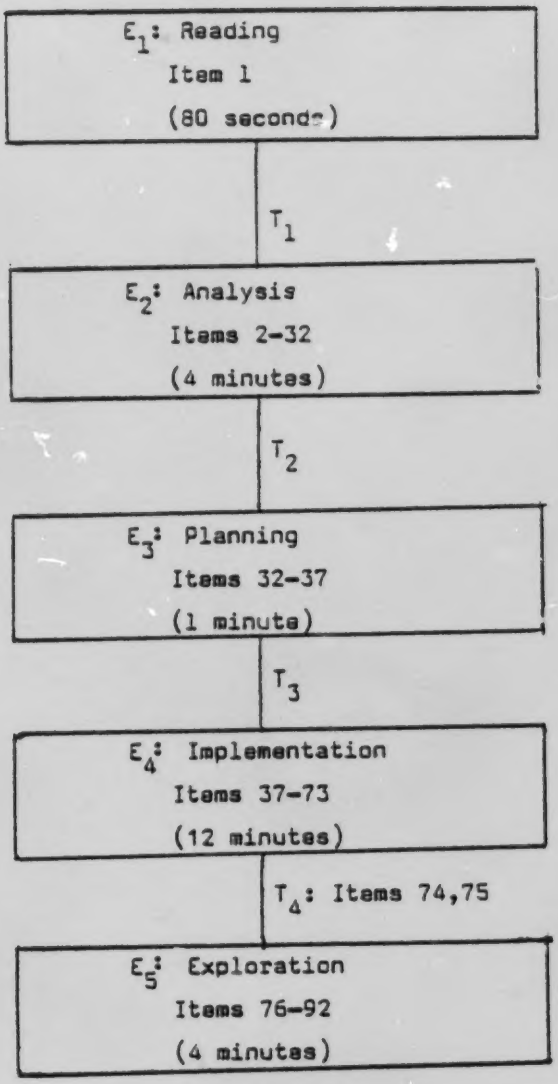


Figure 2

A Parsing of Protocol 2

Detailed analyses of Figures 1 and 2 will not be given, since protocols 1 and 2 have been discussed at some length. (Observe, however, how Figures 1 and 2 reflect the issues singled out for discussion above.) A third protocol will be analyzed in detail.

Both parsing into episodes and delineating "new information" points, turn out to be (more or less) objective decisions. In fact, the parsing of all three protocols that I use in this paper was derived, in consensus, by three undergraduates who followed my instructions but arrived at their characterizations of the protocols in my absence. Reliability in parsing protocols is quite high. (This does not, however, obviate the need for an appropriate formalism: see the final commentary.)

Subjectivity lurks around the corner, however. It is, in fact, already present in the labeling of the episodes given in Figures 1 and 2. This labeling was essential: see the note below.* Any episode is characterized as one

---

*The potential for "combinatorial explosion" in characterizing managerial behaviors is enormous. Managerial behaviors include selecting perspectives and frameworks for a problem; deciding at branch points which direction a solution should take; deciding whether, in the light of new information, a path already embarked upon should be abandoned; deciding what (if anything) should be salvaged from attempts that are abandoned or paths that are not taken; monitoring tactical implementation against a template of expectations for signs that intervention might be appropriate; and much, much more. My early attempts at analyses of managerial behavior called for examining protocols at all managerial decision points and evaluating at each one a series of questions encompassing the issues just mentioned. This approach, while comprehensive, was completely unwieldy. For example, questions about the assessment of state when (a) one has just read the problem, (b) one is "stuck," and (c) a solution has been obtained, are almost mutually exclusive. Thus at any decision point 90% of the questions that might be asked were irrelevant. The framework described above provides a workable compromise.

of the following: Reading, Analysis, Planning, Implementation (or Planning/Implementation if the two are linked), Exploration, Verification, or Transition. What follows is the heart of the analytic framework. There is a brief description of the nature of each type of episode, followed by a series of questions to be asked about each episode once it has been labeled. The parsing, plus the answers to the questions, provide the characterization of the protocol.

Admittedly, these questions are a mixed bag. Some can be answered objectively at the point in the protocol at which they are asked, some in the light of later evidence; some call for inferences or judgments about problem-solving behavior. Further, some ask about the "reasonableness" of certain behavior. Asking questions in this way, of course, begs the significant question: what is a model of "reasonable" behavior? The creation of such models is the crucial long-term question, and there is no attempt to finesse it here. At present, however, we will deal with the notion subjectively, to better understand managerial behaviors so that we can create those models. Though highly subjective, these assessments can be made reliably: agreement between my ratings and the consensus scorings of my students was quite high. To quote Mr. Justice Stewart (1964), "I shall not today attempt to further define the kind of materials I understand to be embraced within that shorthand definition;... But I know it when I see it."

Episodes and the Associated Questions

1. READING.

The reading episode begins when a subject starts to read the problem statement aloud. It includes the ingestion of the problem conditions, and

continues through any silence that may follow the reading--silence that may indicate contemplation of the problem statement, the (non-vocal) rereading of the problem, or blank thoughts. It continues as well through vocal re-readings and verbalizations of parts of the problem statement (observe that in protocol 1, reading included items 1-4).

READING Questions:

a. Have all of the conditions of the problem been noted? (Explicitly or implicitly?)

b. Has the goal state been correctly noted? (Again, explicitly or implicitly?)

c. Is there an assessment of the current state of the problem solver's knowledge relative to the problem-solving task (see TRANSITION)?

2. ANALYSIS.

If there is no apparent way to proceed after the problem has been read (i.e., a solution is not "schema driven"), the next (ideal) phase of a problem solution is analysis. In analysis, an attempt is made to fully understand a problem, to select an appropriate perspective and to reformulate the problem in those terms, and to introduce for consideration whatever principles or mechanisms might be appropriate. The problem may be simplified or reformulated. (Often analysis leads directly into plan development, in which case it serves as a transition. Of course, this episode may be bypassed completely.)

ANALYSIS questions:

a. What choice of perspective is made? Is the choice made explicitly, or by default?

b. Are the actions driven by the conditions of the problem? (working forwards)

c. Are the actions driven by the goals of the problem? (working backwards)

d. Is a relationship between conditions and goals sought?

e. Is the episode, as a whole, coherent? In sum (considering a-d), are the actions reasonable? (comments?)

3. EXPLORATION.

Both its structure and content serve to distinguish exploration from analysis. Analysis is generally well structured, sticking rather closely to the conditions or goals of the problem. Exploration, on the other hand, is less well structured and further removed from the original problem. It is a broad tour through the problem space, a search for relevant information that can be incorporated into the analysis/plan/implementation sequence. (One may well return to analysis with new information gleaned during exploration.)

In the exploration phase of problem solving one may find a variety of problem-solving heuristics, the examination of related problems, the use of analogies, etc. Though amorphously structured, exploration is not, ideally, without structure: there is a loose metric on the problem space, the perceived distance of objects under consideration from the original problem, that should serve to select items for consideration. Precisely because exploration is weakly structured, both local and global assessments are critical here (see transition as well). A wild goose chase, unchecked, can lead to disaster; but so can the dismissal of a promising alternative.

If new information arises during exploration but is not used, or the examination of it is tentative, "fading in and fading out," the coding scheme calls for delineating "new information" within the episode. If, however, the

problem solver decides to abandon one approach and start another, the coding scheme calls for closing the first episode, denoting (and examining) the transition, and opening another exploration episode.

EXPLORATION questions:

a. Is the episode condition driven? Goal driven?

b. Is the action directed or focused? Is it purposeful?

c. Is there any monitoring of progress? What are the consequences for the solution of the presence or absence of such monitoring?

d. At NEW INFORMATION points (including the introduction of heuristics) and LOCAL ASSESSMENT points:

1. Does the problem solver assess the current state of his knowledge? (Was it appropriate??)

2. Does the problem solver assess the relevancy or utility of the new information? (Was it appropriate?)

3. What are the consequences for the solution of the actions (or inactions) described in 1 and 2 above?

4. PLANNING/IMPLEMENTATION.

Since the emphasis here is on managerial questions, detailed issues regarding plan formation will not be addressed: the primary questions of concern here deal with whether or not the plan is well-structured, whether the implementation of the plan is orderly, and whether there is monitoring or assessment of the process on the part of the problem solver(s), with feedback to planning and assessment at local and/or global levels. Many of these judgments are subjective. For example, the absence of any overt planning behavior does not necessarily indicate the absence of a plan; in fact, protocols of

"schema-driven" solutions often proceed directly from the reading episode into the coherent and well structured implementation of a non-verbalized plan. Thus the latitude of the questions below: the scheme should apply to a range of circumstances, from schema-driven solutions to those where the subject happens upon an appropriate plan by design or accident.

PLANNING/IMPLEMENTATION questions:

a. Is there evidence of planning at all? Is the planning overt or must the presence of a plan be inferred from the purposefulness of the subject's behavior?

b. Is the plan relevant to the problem solution? Is it appropriate? Is it well structured?

c. Does the subject assess the quality of the plan as to relevance, appropriateness, or structure? (If so, how do those assessments compare with the judgments in (b)?)

d. Does implementation follow the plan in a structured way?

e. Is there assessment of implementation (especially if things go wrong), at the local or global level?

f. What are the consequences for the solution of assessments if they occur, or if they do not?

5. VERIFICATION.

The nature of the episode itself is obvious.

a. Does the problem solver review the solution?

b. Is the solution tested in any way? (If so, how?)

c. Is there any assessment of the solution, either an evaluation of the process or assessment of confidence in the result?

6. TRANSITION.

The juncture between episodes is, in most cases, where managerial decisions (or their absence) will make or break a solution. Observe, however, that the presence or absence of assessment or other overt managerial behavior cannot necessarily be taken as either good or bad for a solution. In an expert's solution of a routine problem, for example, the only actions one sees may be reading and implementation. This explains, in part, the contorted and subjective nature of what follows.

TRANSITION questions:

a. Is there an assessment of the current solution state, and any attempt to salvage or store things that might be valuable in it?

b. What are the local and global effects on the solution of the presence or absence of assessment in part a? Was the action there appropriate or necessary?

c. Is there an assessment of the short and/or long term effects on the solution of the new direction, or does the subject simply "jump into" the new approach?

d. What are the local and global effects on the solution of the presence or absence of assessment in part c? Was the action there appropriate or necessary?

The Full Analysis of a Protocol

Appendix 3 presents the full protocol of two students working on the following problem:

> Consider the set of all triangles whose perimeter is a fixed number, P. Of these, which has the largest area? Justify your answer as best you can.

Student K is the same student that appeared in protocol 1. Student D (not the same as student D in protocol 2) was a freshmen with one semester of calculus behind him. This protocol was taken at the end of my problem-solving course, while protocols 1 and 2 were taken at the beginning.

The parsing of protocol 3 is given in Figure 3. The analysis given below follows that parsing.

Insert Figure 3 about here

Episode 1 (Reading, items 1, 2)

a. The conditions were noted, explicitly.

b. The goal state was noted, but somewhat carelessly (items 10, 11).

c. There were no assessments, simply a jump into exploration.

Transition 1 (Null)

a, b, c, d. There were no serious assessments of either current knowledge or of directions to come. These might have been costly, but were not--assessments did come in E_2.

Episode 2 (Exploration, items 3-17)

a. The explorations seemed vaguely goal-driven.

b. The actions seemed unfocused.

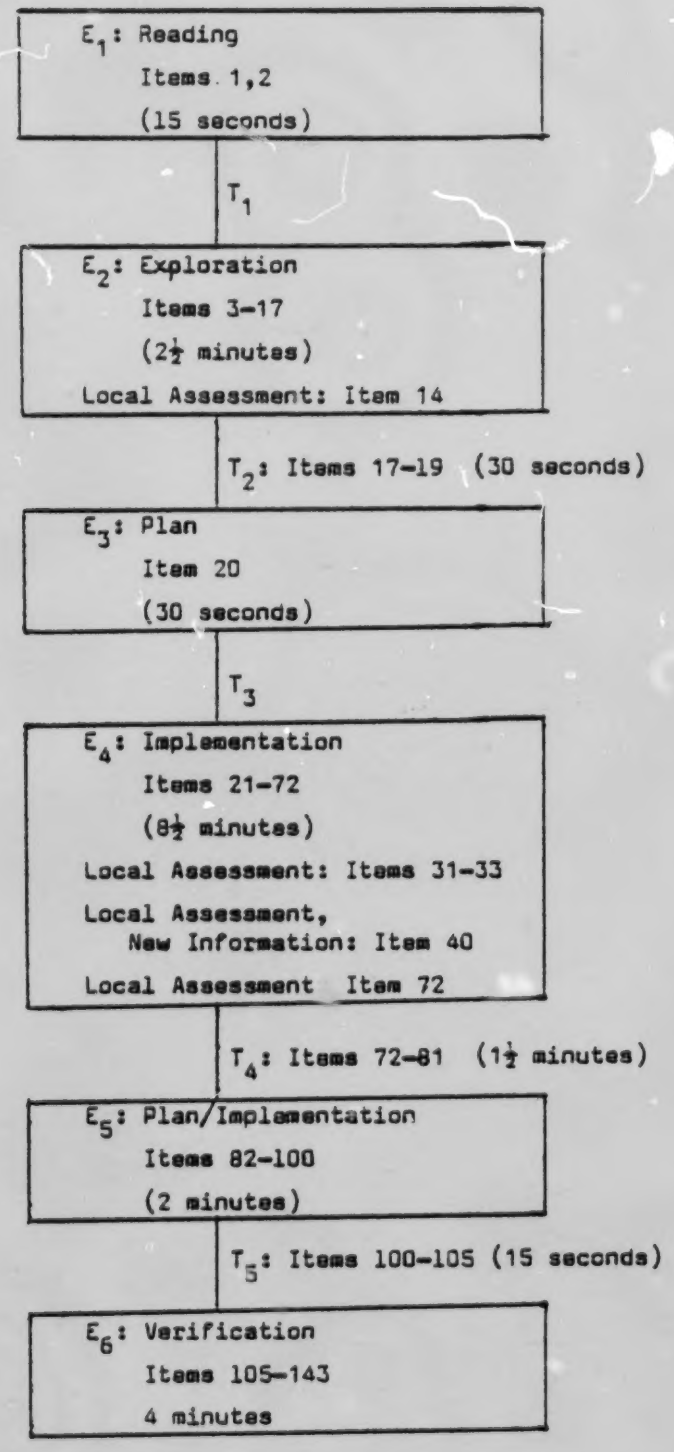


Figure 3

A Parsing of Protocol 3

c, d. There was monitoring, at items 14-17. This grounded the explorations, and led into Transition 2.

Transition 2 (Items 17-19)

a, b, c, d. Assessments were made both of what the students knew, and of the utility of the conjecture they made. The result was the establishment of a major direction: try to prove that the equilateral triangle has the desired property, and of a plan (episode 3). NOTE: If this seems inconsequential, contrast this behavior with the transition T_1 in protocol 1. The lack of assessment there, in virtually identical circumstances, sent the students on a 20 minute wild goose chase!

Episode 3 (Plan, item 20)

a. The plan is overt.

b. It is relevant and well structured. As to appropriateness and assessment, see the discussion of T_3.

Transition 3 (Null)

a, b. There was little of value preceding the plan in item 20; the questions are moot.

c. There was no assessment of the plan; there was immediate implementation.

d. The plan was relevant but only dealt with half of the problem: showing the largest isosceles was the equilateral. The "other half" is to show that the largest triangle must be isosceles, without which this part of the solution is worthless. . .a point realized somewhat in item 72, 8 minutes later. The result was a good deal of wasted effort. The entire solution was not sabotaged, however, because monitoring and feedback mechanisms caused the termination of the implementation episode (see the sequel).

Episode 4 (Implementation, items 21-72)

a. Implementation followed the lines set out in episode 3, albeit in somewhat careless form. The conditions were somewhat muddied as the first differentiation was set up. The next two local assessments corrected for that (better late than never).

Local Assessment (Items 31-33)

1, 2, 3. The physically unrealistic answer caused a closer look at the conditions--but not yet a global reassessment (possibly not called for yet).

Local Assessment, New Information (Item 40)

1, 2, 3. The "new information" here was the realization that one of the problem conditions had been omitted from their implementation ("we don't set any conditions--we're leaving P out of that"). This sent them back to the original plan, without global assessment. The cost: squandered energy until item 72.

Local/Global Assessment (Item 72)

This closes E_4. See T_4.

Transition 4 (Items 72-81)

a, b. The previous episode was abandoned, reasonably. The goal of that episode, "show it's the equilateral," remained. This, too, was reasonable.

c, d. They ease into Episode 5 in item 82. (It's difficult to say how reasonable this is. Had they chosen something that didn't work, it might have been considered meandering. But what they chose did work.)

Episode 5 (Plan/Implementation, items 82-100)

a, b. "Set our base equal to something" is an obviously relevant heuristic.

c. They plunge ahead as usual.

d. The variational argument evolved in a semmingly natural way.

e. There was local assessment (item 95). That led to a rehearsal of the sub-argument (item 96), from which D apparently "saw" the rest of the solution. Further (item 100), D assesses the quality of the solution and his confidence in the result.

Transition 5 (Items 100-105)

a, b, c, d. The sequel is most likely the result of a two-person dialectic. It appears that D was content with his solution (perhaps prematurely), although his clarity in explaining his argument in E_6 suggests he may have been justified.

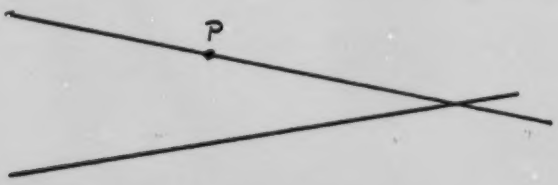
Episode 6 (Verification, items 105-143)

This is not a verification episode in the usual sense. K's unwillingness to rest until he understands forced D into a full rehearsal of the argument and a detailed explanation, the result being that they are both content with the (correct) solution.

Some Empirical Results

Protocols 1 and 2 are relatively typical of the dozen protocols taken from pairs of students (six pairs, two problems for each pair) before a month-long intensive problem-solving course that focused on both tactics (heuristics) and strategies. The first problem was the one discussed in protocols 1 and 2, to find the largest triangle that can be inscribed in a circle. The second problem was a geometric construction:

> You are given two intersecting straight lines, and a point marked on one of them, as on the figure below. Show how to construct, using a straightedge and compass, a circle which is tangent to both lines and has the point P as its point of tangency to one of the lines.



Brief "snapshots" of a few representative pretest protocols are given below. These are too condensed to be useful for model building, but serve to demonstrate again the critical importance of managerial or strategic decision making. They also stand in (partial) contrast to the students' posttest behavior and (stark) contrast to some expert behavior. The diagrams that represent our episode analyses are here condensed into a sequential list of episode titles, with transitions deleted if there were none. Thus Figure 1 is rendered as (Reading/T_1/Exploration), etc.

E.T. & D.R., Problem 1. (Reading/T_1/Exploration)

After a brief mention of "max-min" problems, and a brief caveat ("But will it apply for all cases? I don't know if we can check it afterwards") in transition, they set off to calculate the area of the equilateral triangle. So much for the next fifteen minutes; in spite of some local assessments ("this isn't getting us anywhere") they continued those explorations. Result: all wasted effort.

E.T. & D.R., Problem 2 (Reading/Exploration)

In the initial explorations a series of sketches contains all the vital information they need to solve the problem, but they (without any attempt at review or assessment) overlook it. The solution attempt is undirected and rambling. Possibly because they feel the need to do something, they try their hand at an actual construction--already shown to be incorrect by their sketches--and are stymied when it doesn't work. Overall: lost opportunities, unfocused work, wasted effort.

Note: E.T. and D.R. are both bright; both had just completed the first semester calculus course with A's.

D.K. & B.M., Problem 2 (Read/Analyze/T_1/Explore/Analyze(Solve)/Verify)

Analysis is extended and coherent, but followed by a poor transition into an inappropriate construction that deflects the students off track for three and a half minutes. When this doesn't work they return to analysis and solve the problem. A detailed verification seals things up. Managerial decisions worked reasonably well here.

B.W. & S.H., Problem 2 (Reading/Exploration/T_1/Exploration)

A series of intuition-based conjectures led to a series of attempted

constructions, the last of which happened to be correct--though neither student had any idea why, and they were content that it "looked right." This was a classic trial-and-error tape, and only because the trial space was small was there a chance that the right solution would be hit upon. There was one weak assessment (after a construction) that constituted T_1, but the result was simply a continuation of trial-and-error search.

Impetuous jumps into a particular direction were pretty much the norm in the pretests, and these first approaches were rarely curtailed. (This behavior was so frequent that it earned the name "proof by assumption," coined by my assistants.) Since there was little assessment and curtailment, little was ever salvaged from an incorrect first attempt, and a solution was often doomed to failure in the first few minutes of exploration.

Protocol 3, which has been discussed above, was taken after the problem-solving course. It is a representative, perhaps slightly better than average, sample of post-instruction performance. What makes this tape "better" than pretest tapes is not that the students solved the problem, for their discovery of the variational argument that solves it may have been serendipitous. However, that they had the time to consider the approach was no accident: they had evaluated and curtailed other possible approaches as they worked on the problem. In general there was more evaluation and curtailment on the posttests than on the pretests, and less pursuit of "wild goose chases." In some cases this allowed for a solution, in some not; but at least their actions did not preclude the possibility. The following statistic summarizes the difference:

Seven of the twelve pretest protocols were of the type (Reading/Exploration);

Only two of the twelve posttest protocols were of that type. Not at all coincidentally, their performance improved on a variety of other measures as well (Schoenfeld, Note 7). However, the overall quality of the students' managerial monitoring, assessing, and decision making on the posttests was still quite poor. To indicate the contrast in managerial behaviors between experts and novices, we turn to the protocol of an expert working on a geometry problem. The expert, a number theorist, had a broad mathematical background but had not dealt with geometric problems for a number of years. It shows. By some standards, his solution is clumsy and inelegant. (In a department meeting it was held up for ridicule by the colleague who produced Protocol 5.) Precisely because the expert does run into problems, however, we have the opportunity to see the impact of his metacognitive, managerial skills.

The episode analysis of Protocol 4 is given in Figure 4. For (obvious) reasons of space, the full analysis will be condensed.

Insert Figure 4 about here

The critical point to observe in this protocol is that a monitor/assessor/manager is always close at hand during the solution. Rarely does more than a minute pass without some clear indication that the entire solution process is being watched and controlled, both at the local and global levels. The initial actions are an attempt to fully understand the given problem. By item 3 there is the awareness that some other information, or observation, will be necessary in order for a solution to be obtained. The actions in items 4 and 5 are goal-driven and, in item 6, yield the necessary information. This is utilized immediately in items 7-8. There is a (meta) comment that the first part of the

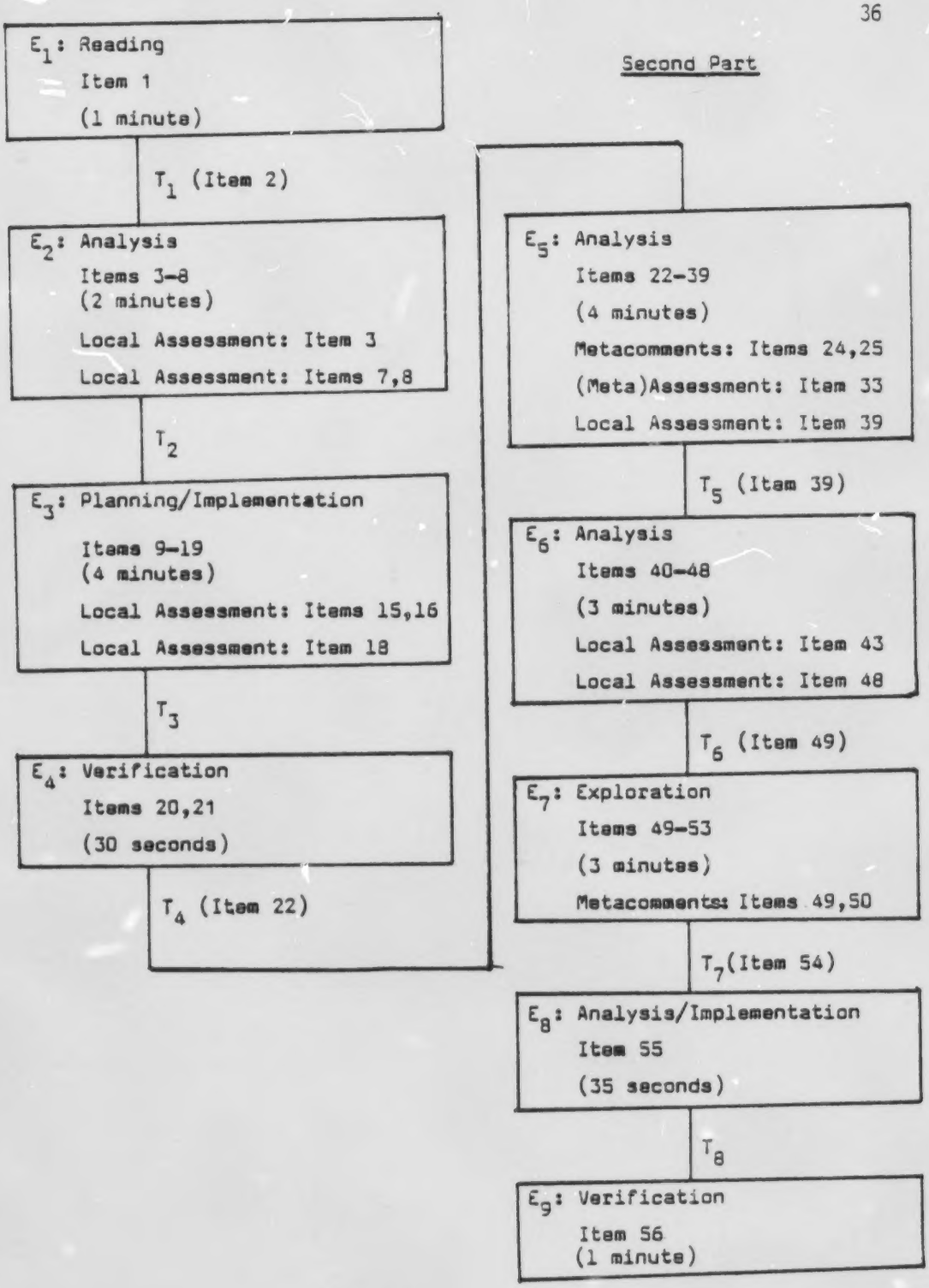


Figure 4

A Parsing of Protocol 4

problem will be solved with one construction, which can be made. The plan is made in item 9. Implementation is interrupted twice with refinements (items 15 & 16; item 18) that again indicate that the subject is on guard for clarifications and simplifications at almost all times. The first part of the problem concludes with a quick but adequate rehearsal of the argument.

Like part 1, the second part of the solution begins with a qualitative analysis of the problem. In item 24, there is a comment that "this is going to be interesting" (i.e., difficult). Such a preliminary assessment of difficulty is, I believe, an indication of an important element of experts' metacognitive behavior. Experts seem to judge their work against a "template of expectations" when solving a problem. These expectations may be major factors in the experts' decisions to pursue or curtail various lines of exploration during the problem-solving process.

The solution of the second part continues, well structured, with a coherent attempt to narrow down the number of cases that must be considered. This is an implementation of "that kind of induction thought" from item 29. It appears to be a "forward" or "positive" derivation, verifying that all of the cases can be done. Yet the phrase "no contradiction" in item 33 reveals that the problem solver retains an open mind about whether the constructions could actually be implemented, and is still probing for trouble spots. The potential for a reversal, using argument by contradiction if he should come to believe one of the constructions impossible, is very close to the surface. This distanced overview, and the maintenance of a somewhat impartial perspective, are confirmed in item 49.

Assessment is, likewise, always in the immediate vicinity. The comment "if so this can be done in one shot," in item 40 indicates not only that solutions

are planned ahead, but that the plans are assessed. Even the rather unusual excursion into quadratic extensions (item 53) is preceded by a comment about "knocking this off with a sledgehammer," and quickly curtailed.

In sum: this rather clumsy solution (see Protocol 5 in contrast), with its apparent meandering through the solution space, is in reality rather closely controlled. There is constant monitoring of the solution process, both at the tactical and strategic levels. Plans and their implementation are continually assessed, and acted upon in accordance with the assessments. Tactical, subject-matter knowledge plays a minor role here: metacognitive, "managerial" skills provide the key to success.

Discussion

This paper raises many more questions than it can answer. It was intended to. The extended discussions of protocols were designed to make one point absolutely clear: "metacognitive" or "managerial" skills are of paramount importance in human problem solving. As Brown observed (1978, p. 82), these types of decisions "are perhaps the crux of intelligent problem solving because the use of an appropriate piece of knowledge...at the right time and in the right place is the essence of intelligence." The inverse of this proposition should be given comparable stress: avoiding inappropriate strategies or tactics, at the wrong time or in the wrong place, is an equally strong component of intelligent problem solving.

To deal coherently with such executive decision making, one needs a framework for examining, modeling, and judging it. This kind of framework must, perforce, be substantially different from extant schemes like those used in mathematics education (Lucas, et al., 1979; Kantowski, Note 3), that focus on overt behaviors at a detailed level. As we saw in Protocol 1, the absence of an

assessment may doom an entire solution to failure. Schemes that only seek overt behaviors cannot hope to adequately explain that protocol.

This kind of framework must also differ substantially from those used in Artificial Intelligence to simulate expert behavior in areas such as physics. Larkin, et al., (1980) characterize such work as depending on production systems to simulate the pattern recognition that "guide[s] the expert in a fraction of a second to relevant parts of the knowledge store...[and] guide[s] a problem's interpretation and solution (p. 1336)." While aspects of Protocol 4 such as the recognition of similar triangles (item 6) are compatible with this perspective, the whole of Protocol 4 stands in sharp opposition to it. At least half of the action in that protocol is metacognitive; it almost seems as if "manager" and "implementer" work in partnership to solve the problem. And it is precisely when the expert's problem-solving schemata (or "productions") do not work well that the managerial skills serve to constitute expertise.

The framework presented in this paper provides a mechanism for focusing directly on certain kinds of managerial decisions. Since a manager ought to be present at major turning points in a problem solution (if only to watch, in case action is necessary), the transition points between "episodes" are the logical place to look for the presence, or absence, of such decision making. Here we come to the first serious question: what, precisely, constitutes an "episode"? While there is reliability among coders in parsing these protocols at the macroscopic level, that begs the question: we need a rigorous formalism for characterizing such episodes. Unfortunately, I have not been able to adapt schemata for story understanding or for episodes in memory (see Bobrow and Collins, 1975) to deal with these kinds of macroscopic problem-solving episodes.

A formalism needs to be developed.

Questions regarding the characterization and evaluation of the monitoring, assessing, and decision making processes during problem solving are far more thorny. The role of the monitor was quite clear in Protocol 4; it assured that the solution stayed "on track." But how are these decisions made? It is clear from a variety of expert protocols that a priori expectations of problem or subtask difficulty serve as a basis for the decision to intervene. But the nature of the monitoring, the criteria for assessments, what the tolerances are, and how intervention is triggered all remain to be elaborated.

Similarly, assessment is not always desirable or appropriate: in a schema-driven solution, for example, one should simply implement the solution unless or until something untoward pops up. A simple-minded model that looked for assessment at each transition point between episodes (and other places) would miss the point entirely: assessment is only valuable some of the time, and we need to know when (and how).

In the long run, we need a detailed model of managerial monitoring, and assessment, and of the criteria used for assessment and decision making. This model will enable us to answer questions like those for the transition phase, "was the action or inaction appropriate or necessary?" In the meantime, these questions are not an evasion: they are an attempt to gather data so that the model can be constructed. A further refinement of these questions, and a much more detailed characterization of metacognitive acts in general, will be necessary. I hope that this paper provides a step in that direction.

Reference Notes

1. Ericsson, K. A. & Simon, H. A. Retrospective verbal reports as data. C.I.P. Working paper No. 388, 1978.

2. Ericsson, K. A. & Simon, H. A. Thinking-aloud protocols as data. C.I.P. Working paper No. 397, 1979.

3. Kantowski, M. G. The use of heuristics in problem solving: an exploratory study (Appendix D, Final Tech. Rep. NSF project SED 77-18543).

4. Kulm, G. Personal communication, "Process Code Dictionary," revised December 22, 1979.

5. Neches, R. Promoting self-discovery of improved strategies. C.I.P. Working paper No. 398, 1979.

6. Newell, A. On the analysis of human problem solving protocols. Paper presented at the International Symposium on Mathematical and Computational Methods in the Social Sciences, Rome, July 1966.

7. Schoenfeld, A. Measures of Problem Solving Performance and of Problem Solving Instruction. Manuscript submitted for publication, 1980.

References

Anzai, Y. & Simon, H. A. The theory of learning by doing. Psychological Review, 1979, 86(2), 124-140.

Bobrow, D. & Collins, A. Representation and Understanding. New York: Academic Press, 1975.

Brown, A. Knowing when, where, and how to remember: a problem of metacognition. In R. Glaser (Ed.), Advances in Instructional Psychology (Vol. 1). Hillsdale, New Jersey: Lawrence Earlbaum Associates, 1978.

Larkin, J. McDermott, J., Simon, D., Simon, H. Expert and novice performance in solving physics problems. Science, 20 June 1980, 208, 1335-1342.

Lucas, J. F., Branca, N., Goldberg, D., Kantowski, M. G., Kellogg, H., & Smith, J. P. A process-sequence coding system for behavioral analysis of mathematical problem solving. Columbus, Ohio: Ohio State University, 1979. (ERIC Mathematics Education Information Report Task variables in mathematical problem solving)

McDermott, J. & Forgy, C. Production system conflict resolution strategies. In D. A. Waterman & F. Hayes (Eds.), Pattern-directed inference systems. New York: Academic Press, 1978.

Neves, D. M. A computer program that learns algebraic procedures by examining examples and by working test problems in a textbook. In Proceedings of the second annual conference of the Canadian Society for Computational Studies of Intelligence, Toronto, 1978.

Polya, G. Mathematical discovery (Vol. 2). New York: Wiley, 1965.

Schoenfeld, A. H. Teaching problem solving in college mathematics: the elements of a theory and a report on the teaching of general mathematical problem-solving skills. In R. Lesh, D. Mierkiewicz & M. Kantowski (Eds.), Applied mathematical problem solving. Columbus, Ohio: Ohio State University, 1979 (ERIC Document).

Schoenfeld, A. H. Teaching problem-solving skills. American Mathematical Monthly, 1980, 87(10), 794-805.

Simon, H. Problem solving and education. In D. T. Tuma & F. Reif (Eds.), Problem solving and education: issues in teaching and research. New York: Lawrence Erlbaum Associates, 1980.

Stewart, P. Jacobellis v. Ohio. Decision of United States Supreme Court, 1964.

Protocol 1

1. K: (Reads roblem) Three points are chosen on the circumference of a circle of radius R, and the triangle containing them is drawn. What choice of points results in the triangle with the largest possible area? Justify your answer as best as you can.

   You can't have an area larger than the circle. So, you can start by saying that the area is less than $1/2\pi R^2$.

2. A: O.k. So we have sort of circle--3 points in front and R here and we have let's see--points--

3. K: We want the largest one--

4. K: We want the largest one--

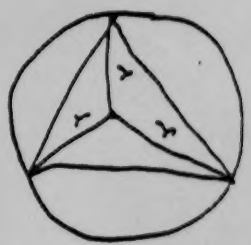
5. A: Right, I think the largest triangle should probably be equilateral. O.k., and the area couldn't be larger than πR^2.

6. K: So we have to divide the circumference of the three equal arcs to get this length here. That's true. Right. So, 60-120 arc degrees--o.k.--so, let's see, say that it equals R over S--this radius doesn't help.

7. A: Do we have to justify your answer as best as you can? Justify why this triangle-----justify why you-----o.k. Right.

8. K: O.k. Let's somehow take a right triangle and see what we get. We'll get a right angle.

9. A: Center of circle of right triangle. Let's just see what a right triangle--is this point in the center? Yep, o.k. Yeah.



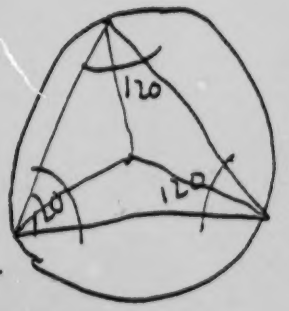
10. K: This must be the radius and we'll figure out that'll be like that, right?

11. A: So the area of this--

12. K: is R, is R--1/2 base times height, that's S and 2R, height is R so it is $1/2R^2$. It's off by a factor of 2.

13. A: O.k. But what we'll need is to say things like--o.k. Let's go back to the angle--probably we can do something with the angle.

14. K: Oh, I got it! Here, this is going to be 120--the angle of 120 up here--



15. A: Right! Yes, this is 120 and this is 120.

16. K: Right!

17. A: So--

18. K: We have to figure out--

19. A: Why do we choose 120--because it is the biggest area--we just give the between the biggest area--120.

20. K: Umm. Well--the base and height will be equal at all times.

21. A: Base and height--right--

22. K: In other words--every right triangle will be the same.

23. A: Ah, ah--we have to try to use R, too.

24. K: Right.

25. A: O.k. (seems to reread problem)--justify your answer as best as you can. O.k. (pause)

26. A: So--there is the picture again, right? This is--both sides are equal--at this point--equal arc, equal angles--equal sides--this must be the center and this is the radius R--this is the radius R--

27. K: So we have divided a triangle with three equal parts and--

28. A: There used to be a problem--I don't know about something being square--the square being the biggest part of the area--do you remember anything about it?

29. K: No..I agree with you--the largest area...of something in a circle, maybe a rectangle, something like that...

30. A: Oh, well...so...

31. K: Since this is R--and this is going to be 120, wouldn't these two be R also?

32. A: Right.

33. K: This is 120.

34. A: Ah, ah.

35. K: Like a similar triangle--120 and 120 are the same angle--so these two should be R.

36. A: O.k. Maybe they are.

37. K: Why can't they be?

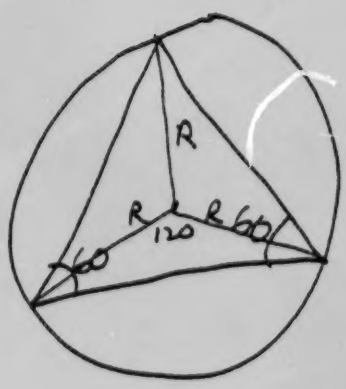
38. A: Mumbles----------

39. K: See, look--this is the angle of 120--right?

40. A: Right.

41. K: And this is an angle of 120. Right? This is like similar triangles--

42. A: Wait a second--I think if you--this is true 120 but I don't think this one is-------



43. K: It is an equilateral triangle--that's--

44. A: No--it should be a 60.

45. K: That's right--it should be a 60. Mumbles------that's 1/2 of it---that's right--2R.

46. A: What are you trying to read from?

47. K: What if we could get one of these sides, we could figure out the whole area.

48. A: Ah, ah.

49. K: Right?

50. A: Presume this to be 1/2 that side, we've got 1/2 base times height. We'll get the area--all we have to show is the biggest one.

51. K: When we take the formula πR^2, minus 1/2 base times height and then maximize that--then take the derivative and set it equal to zero. We can get that function--then we can get this in the form of R.

52. A: O.k.

53. K: Then we can try this as the largest area.

54. A: Do you want to get this function, this as a function of R?

55. K: Yeah.

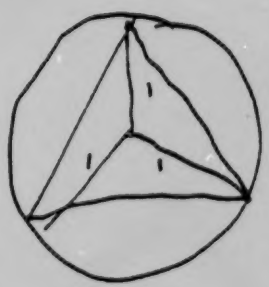
56. A: We can, I think. So you want this--right?

57. K: Well, it is kind of obvious that with B & H you are still going to have an R in it. So you can subtract it.

58. A: You have H in it. Well we have this one here. Mumbles--- (repeats the problem). Try this to be 2R.

59. K: No--it can't be. It has to be between R and 2R.

60. A: Yeah.



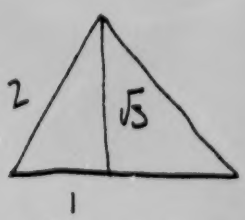
61. K: Helps us a lot! Set R equal to 1.

62. A: R = 1?

63. K: Right.

64. A: O.k.

65. K: That's one, that's one, that's one--it'll equal S over R. The area of the triangle is equal with R = 1, it's 2.



66. A: Well...height equals...

67. K: That's for the sides of the triangle--that's obvious -R = 1.

68. A: O.k.--divided into equal parts---(lots of mumbling)--This from---well--you know--o.k. If you see we probably try to fix one point and choose the other two--o.k.--we are going to go from something that looks like this all the way down---

69. K: Right.

70. A: Right. O.k. and here the height is increasing where the base is decreasing.

71. K: Right. (Mumbles)

72. A: When we reach----o.k.

73. K: What is the area, side squared over 4 radical 2 for

an equilateral triangle? Is it like that?

74. A: You want the area for an equilateral triangle.

75. K: The area? I don't know. Something like side squared over radical 2, or something--

76. A: If you can probably show...at a certain point where we have the equilateral triangle the base and the...well...you know the product of the base since the base is decreasing and the height is increasing every time we move the line. If you can show a certain point, this product is the maximum--so we have the area is a maximum at that point. So this one is decreasing-----And at this point we have R, R, and R.

77. K: Ah, ah.

78. A: O.k. This is the base--is 2R--a right angle.

79. K: It wouldn't be $2R^2$.

80. A: Mumbles----One more--I mean--

81. K: O.k.

82. A: It should be R^2. But base times height--mumbles--and this one, say this is R + X.

83. K: The height equals R + X, so the base equals R-X.

84. A: Mumbles--those two things are equal to this--

85. K: Right.

86. A: All right.

87. K: I don't know.

88. A: We want this product of h as a maximum--as a maximum--and this one...I don't know.

Appendix 2

Protocol 2

1. D: Reads the question.

2. B: Do we need calculus for this? So we can minimize, or rather maximize it.

3. D: My guess would be more like--mumbling--my basic hunch would be that it would be--

4. B: An equilateral--

5. D: 60, 60, 60.

6. B: Yeah.

7. D: So what choice of points has to be where on the triangle--these points are gonna be.

8. B: Try doing it with calculus--see if you can--just draw the circle--see what we'll do is figure out the right triangle--

9. D: Yeah,or why don't we find--or why don't we know the--some way to break this problem down into--like what would a triangle be for half the circle?

10. B: 60 degrees here?

11. D: Why don't we, why don't we say that--o.k.--why don't we find the largest triangle with base--one of the diameters, o.k.

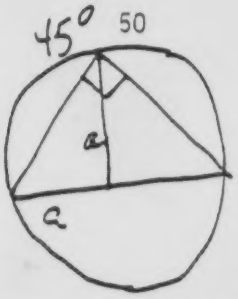
12. B: Base as one of the diameters?

13. D: Yeah.

14. B: O.k. That would be just a family of right triangles--that go like this.

15. D: And they're all the same area?

16. B: No, no they're not all the same area--the biggest area would be in one like that. See if we could figure out--make it into sort of like a--if we could do it with calculus and I know there is a way. I just don't remember how to do it.



17. D: I have a feeling we wouldn't need the calculus. So this area then this is r and this would be--r^2--that would be the area of this--so then the distance here has got to be--45 degrees--

18. B: Right--that's got to be 45 degrees because they are the same. That's A--A over square root of 2--right?

19. D: Umma.

20. B: If that's radius--A--and this is A, too, so that would be A^2, that would be r^2, wouldn't it?

$0 \leq \theta < 45^\circ$

21. D: Right.

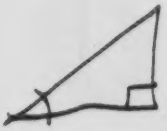
22. B: But I think this would be bigger.

23. D: Oh, of course it would be bigger--I was just wondering if... (Pause)

24. D: Well we can't build a diamond--so we can't build a diamond that would go like that, obviously you want to make it perfectly symmetrical, but we can, if we maximize this area, and just flip it over, if we can assume that it is going to be symmetrical.

25. B: Yeah, it is symmetrical.

26. D: And if we can find the best area--



27. B: You mean the best--cut it in half in a semicircle.

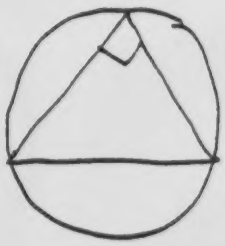
28. D: Right. And if we can find the best area of--

29. B: Any triangle that fits in a semicircle--well it wouldn't be a semi-

30. D: No it's a semicircle.

31. B: Largest triangle that fits in there?

32. D: Yeah, but it would have to be--if it is going to be symmetrical though, then you know this line has to be flat--it is going to have to form a right angle. So all we really have to do is form a right angle. So all we really have to do is find the largest area of a right triangle--inscribed in a semicircle.



33. B: Largest area of a right triangle. Yea, but obviously it is this one which is wrong.

34. D: No--No--

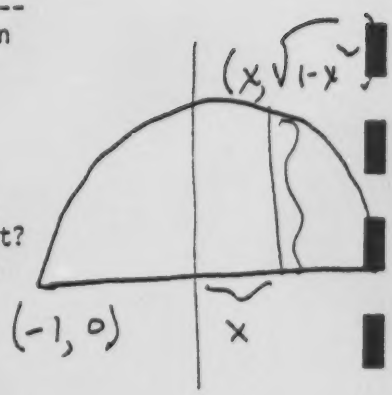
35. B: One like this.

36. D: Yeah with that angle, right.

37. B: O.k.--how we go about doing that? Hey, like we can--use the unit circle, right?

38. D: Umma.

39. B: So that means--this is $(1-x^2)$--this point right here--will be $(1-x^2)$, o.k. this squared--mumbling--I'll just put some points down to see if...pick an arbitrary--



40. D: Yeah, yeah, just to find this point--

41. B: All right, this is 1. Now I've got to find that point--o.k. What is the area of this--this is the distance right here times that distance, right? Product of those distances--area equals from this distance would be this, would be x value which would be x-1 or x+1? O.k., it's x+1, this distance right here times this distance right there which would be the y coordinate which is x^2. Want to take the derivative of that--to the x--mumbling.

$$A = \frac{1}{2}(x+1)\sqrt{1-x^2}$$

42. D: O.k.

43. B: Times (2-x). Did I have, oh, the two is crossed out so I just have an -x--or, that was over $\sqrt{1-x^2}$, plus all this stuff. And set that equal to zero and you get that--oh, this is just one, isn't it--this is just one--so one of that, plus that equals zero, right?

$$\frac{dA}{dx} = (x+1)\left(\frac{1}{2}\right)\left(\frac{-2x}{\sqrt{1-x^2}}\right) + \sqrt{1-x^2}$$

44. D: I think we're getting a little lost here--I am not sure. Well, you go ahead with that--

45. B: Well, I'll just think about it, as it is just mechanical. There is a minus in here, isn't there? Mumbling--o.k. x equals $\sqrt{2}$ and what was this distance, we said? That was x--so that means it would be $\sqrt{2}$--plus 1--that's impossible.

$$\frac{1}{\sqrt{1-x^2}} = \frac{\sqrt{1-x^2}}{1}$$

46. D: Times R.

47. B: If x equals plus or minus the $\sqrt{2}$--

48. D: Umma--

49. B: This y thing would be 1 minus x^2, right?

50. D: This is just the distance--therefore, this right here has to be $\sqrt{2}$. Guess your calculations are all right.

51. B: Yeah,if I got x equals square root of 2--we've got a semicircle here, right? O.k.--and I have the points--right, it's a unit circle and I said that $x^2+y^2 = 1$, so $y = \sqrt{1-x^2}$. O.k.? And--(pause)--the x can't equal the square of the two because it would be out there. I know this has to be right but--

52. D: But all kinds of--let's see--well we know already, o.k. that the triangle is not 45, 45, because that would make it too small. O.k.?

53. B: Um--

54. D: So we know this angle is greater than zero and less than 90 degrees--

55. B: I just want to make sure I didn't--so this is x+1, x+1n...and cross multiply to set $1-x^2 = 1$ which means $x = \sqrt{2}$.

56. D: No, it has to be a 60, 60, 60--right triangle--no I am sorry not a right triangle--has to be a 60, 60, 60 triangle--because no matter where you move these vertices, it has to be a 60, 60, 60 triangle--because no matter where you move these vertices--

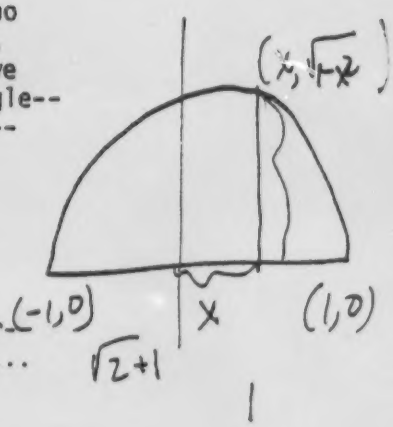
57. B: O.k.

58. D: --you are going to add area to this--like the--mumbling--you are going to add area to this.

59. B: All right, o.k. I understand, but I don't understand why it didn't work for this. I mean that... is there no solution for this equation?

60. D: I don't know--are you sure what you are looking for in that one?

61. B: Yeah,I marked off these and I just wanted to mark off these dimensions.



62. D: O.k. What were you looking for? The length of this?

63. B: I was just looking for the maximum area of this--I said $A = (x+1)\sqrt{1-x^2}$. That's this height which is the square root of $(1-x)^2$. This is the unit circle. That's this distance right here--this minus the x value that I used--x value that is just x. O.k.--cause it is all in terms of x--x minus the x value here, which is x-1, which x+1--so area--ah shoot--I should have put 1/2 that is well,--mumbling--I'll get it. That should be 1/2 there, but I don't think that makes any difference--so that's all in terms of 1.

64. D: So--if--

65. B: Oh, wait a minute there's a difference--so one for two is 1/2 the first part--

66. D: So if you find the maximum area equal to--

67. B: It doesn't make any difference--it is just a factor of 1/2 here--because the area equals 1/2 that.

68. D: No--what's the next move?

69. B: See I get x--see I get a value of x with a plus or minus $\sqrt{2}$, right?

70. D: Umma.

71. B: If I plug x back into this I get $\sqrt{2+1}$, right? Then I plug x back into there and I get $(1-\sqrt{2})^2$ which is $\sqrt{1-1}$ which doesn't work.

72. D: Umma.

73. B: Which doesn't seem right. Plus r^2--mumbling--Let me just check my derivative over again. Now I know my mistake--hold it. I added this x--it's supposed to be times so we've still got a chance. So let me go from there. It is just a derivative mistake. Let me see it will be $(1-x^2)$--no it will be--$(-x+1)$. This might work--if it does--we solve that and cross out this minus 1. That means $x+1+x^2-1$, that makes x^2+x---cross this out--mumbling-- all right? It still doesn't work.

74. D: Well, let's leave the numbers for a while and see if we can do this geometrically.

75. B: Yeah, you're probably right.

76. D: Well, we know that these two are some kind of symmetry.

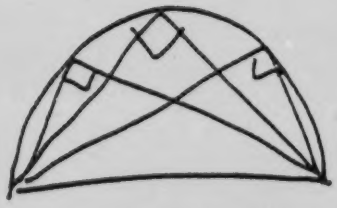
77. B: Yeah.

78. D: I still say we should try--yeah--what we were doing before--just try to fix two of the points and let the third one wander around.

79. B: Yeah, we were going to fix them--yeah, I know what happens if you fix them on the diameter--then you have a family of right triangles.

80. D: Those the maximums.

81. B: Well, I don't see how--where are you going to fix the two points?



82. D: Well, you just fix them on any diameter. You find the largest triangle.

83. B: That would--obviously that would be the 45, 45 triangle if you fix them on the diameter. If you fix them on any chord.

84. D: Yeah, why though. Well, we know that if we put two of the points too close together--o.k.--o.k. --no matter where we put the third point--

85. B: Yeah.

86. D: --it's going to be too small. O.k. If we put them too far apart--o.k.--no matter where we put the third point, we are only using half a triangle.

87. B: O.k.

88. D: So it's got to be--o.k. So--two of the points, at least, well, matter of fact if you've got three points, each two of the points have to be between zero and 1/2 of the circle distance away from each other.

89. B: O.k.

90. D: See how I got that? O.k. so therefore each two of the points has to be like that--so how can we construct a circle that's like

that? O.k. so we stick one point here--arbitrarily--so now the second point has to be somewhere o.k.--within--o.k. in other words, it can't be right here--it can't be right here--it can be anywhere else. We've got to place it so that the third point is going to be within half--

91. B: Half of what--I don't get you there.

92. D: O.k. Now wait a minute--let's see. You know when I said that--(pause). O.k. in other words the relationship between every pair of the three points....

At this point the interviewer (I) terminated the session and asked the students to sum up what they had done. B focused on the algebraic computations he had done in trying to differentiate $(1+x)\sqrt{1-x^2}$. The following dialogue ensued:

I: So what do you wind up doing, when you do that? You wind up finding the area of the largest right triangle that can be inscribed in a semicircle.

D: We determined that.

I. My question is: how does that relate to the original problem?

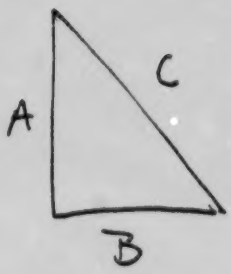
B: Well,...

Appendix 3

Protocol 3

1. K: (Reads problem.) Consider the set of all triangles whose perimeter is a fixed number, P. Of these, which has the largest area? Justify your assertion as best you can. All right now what do we do?

2. D: We got a triangle--well we know we label sides A, B and C.



3. K: Right. I'll make it a right triangle--all right--A,B, C and the relationship such as that 1/2AB = Area and A+B+C = P and $A^2 + B^2 = C^2$ and somehow you've got an area of one of these in the perimeter.

4. D: Yeah,except for somehow--I mean I don't really know--but I doubt that's the triangle of minimum area--well, o.k. we'll try it.

$\frac{1}{2}AB = A$

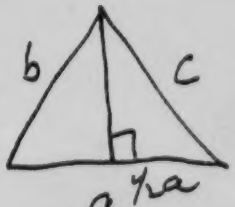
$A+B+C=P$

$A^2+B^2=C^2$

5. K: Largest area. Well, it is the only way we can figure out the area.

6. D: All right.

7. K: But for an isosceles we can do almost the same thing. This ~~is 1/2(A).~~ So that we know that the area is $(A/2)\sqrt{C^2-(A/2)^2}$. The perimeter = A + B + C and the height equals $\sqrt{C^2-(A/2)^2}$.



$b=c$

8. D: All right.

9. K: Now what do we do. We've got to figure out the largest area.

10. D: Isn't it the minimum?

11. K: The largest area.

$A=\frac{1}{2}(a)\sqrt{c^2-(\frac{1}{2}a)^2}$

12. D: So actually if we can get A--we have to get everything in terms of one variable and take the derivative, right? Basically?

$a+b+c=P$

13. K: Yeah,well--

$h=\sqrt{c^2-(\frac{a}{2})^2}$

14. D: Well, I still don't know if we should do--I mean we can find an area for this and can find an area for that, granted, but if we

ever come to a problem like this--I mean we don't know--we have no idea as of yet with a given perimeter what's going to be that.

15. K: Right.

16. D: So, there--I mean--you can do that again but then what do you do?

17. K: Then we're stuck, right? Usually, you know, you could probably take a guess as to what kind of triangle it would be--like you could say it is a right triangle or an isosceles--I think it is an equilateral, but I don't know how to prove it.

18. D: Umma.

19. K: So we have to figure out some way to try to prove that.

20. D: All right, a good guess is that it is an equilateral, then why don't we try an isosceles and if we can find that these two sides have to be equal to form the maximum area, then we can find that--then we should be able to prove that side also has to be equal.

21. K: O.k. so B will be equal to C, so the perimeter P = A + 2B, or A + 2C = P.

22. D: All right.

23. K: Ummmm.

24. D: See what we've got.

25. K: Fix A as a constant then we can do this, solve that for C.

26. D: All right.

27. K: For a maximum area we've got 1/2, let's say A = 1, $)C^2 - 1/4$, right? Maximum area: $1/2(C^2-1/4)^{1/2} = 0$.

28. D: C^2 minus what?

29. K: $(1/2)^2$, yeah, $(1/2)^2$. A/2, where A = 1. O.k.?

$1/2\sqrt{C^2-1/4}$

$1/2(C^2-1/4)^{1/2}=0$

$1/4(C^2-1/4)^{-1/2}(2C)=0$

$2C=0$

$C=0$

30. D: Ah, ah.

31. K: Mumbling--this is $1/4(C^2-1/4)^{-1/2}$. 2C, so we know that 2C has to = 0 and C = 0 and we are stuck!

32. D: We should have taken a derivative in it and everything, you think?

33. K: Yeah, that's the derivative of that. So does it help us? My calculus doesn't seem to work anymore.

34. D: The thing is--pause--you are letting C be the variable, holding A constant. So what was your formula--1/2 base times square root.

35. K: The base A times the square root times the height which is a right triangle to an isosceles which is --so it is $C^2-(A/2)^2$ which would give you this height.

36. D: $A^{2/4}$, no, $A^{2/2}$, no, $(A/2)^2$.

37. K: How about P = , ... no, C = P -A/2? Should we try that--

38. D: No, see part of the thing is, I think that for here we're just saying we have a triangle, an isosceles triangle, what is going to be the largest area? Largest area.

39. K: Largest area--set its derivative equal to 0.

40. D: All right. Well the largest area or the smallest area--I mean--if we are going to take a derivative--I mean--what's going to happen is you have a base and it's going to go down like that--I mean--we don't set any conditions--we're leaving P out of that.

41. K: Ah, ah.

42. D: That's absolutely what we have to stick in.

43. K: We've got C and a P-A over 2.

44. D: P -A over 2.

45. K: Formula--isosceles.

46. D: A + 2B = P--all right?

47. K: Shall we try that--mumbling. -A over 2--we've got to have a minus 1/4 PA--

48. D: Well, then you can put A back in--then you can have everything in terms of A, right? Using this formula, we have the area and we have a--

49. K: All right--P--so that's $A/2\left(\frac{P^2-2A+A^2}{4}-\frac{A^2}{4}\right)^{1/2}$ and that's $A/2\left(\frac{P^2-2A}{4}\right)^{1/2}$...(mumbling and figuring)

50. D: Wait a minute--you just took the derivative of this right here?

51. K: This times the derivative of this plus this times the derivative of this.

52. D: Oh.

53. K: Mumbling and figuring...$A/4\left(\frac{P^2-2A}{4}\right)^{-1/2}(2P-2)+\left(\frac{P^2-2A}{4}\right)^{1/2}$ $1/2 = 0$...so $\frac{2AP-2A}{4}+\frac{P^2-2A}{8}=0$.

54. D: So can we get A in terms of P?

55. K: P^2--

56. D: $8P^2 - 8P^2$ bring the P^2 on this side and multiply it by 8 and we'll have a qudratic in terms--no we won't--then we can just have A we can factor out in the equation--you see.

57. K: O.k. $P^2 =$

58. D: $-8P^2$--oh, are we going to bring everything else to the other side?

59. K: Yeah, 2A- +4A - 4AP x 8--No--

60. D: That's not right. Well, the 8 we can just multiply--

61. K: P^2 = all this.

62. D: Right.

63. K: $P^2 - 4AP =$ --this isn't getting us anywhere.

64. D: $P^2 =$ factor out the A--then we can get A in terms of P.

65. K: $P^2 = 2A$--so you've got $A = \frac{P^2}{6+4P}$ --

66. D: So if we have an isosceles triangle and A has = to--

67. K: be equal to that--

68. D: And if A has to be equal to that and B and C are equal--

69. K: So, B = --(whistles)

70. D: B = P- that.

71. K: 2B = P-A over 2.

72. D: No we aren't getting anything here--we're just getting--thing is that we assumed B to be equal to C so of course, I mean--that doesn't--we want to find out if B is going to be equal to C and we have a certain base--let's start all over, and forget about this. All right, another triangle. Certain altitude.

73. K: Well, let's try to assume that it is an equilateral.

74. D: All right.

75. K: Sides--mumbling--perimeter equals 3S, right?

76. D: Yeah, but wait a minute--that's still not going to really help us--what are we going to do--simply assume that it is an equilateral. We're just going to get that it is an equilateral, of course it is going to be an equilateral if we assume that.

77. K: True.

78. D: We want to prove that it is an equilateral if we think it is. If we want to do anything we can--

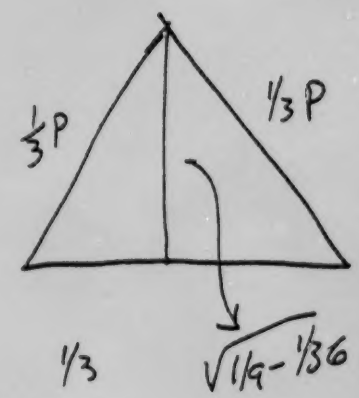
79. K: Yeah, how do you prove it?

80. D: Well, we can make up a perimeter--we don't need a perimeter P, do we? So,--

81. K: Where are you going to get area formula in the form of P?

82. D: We want to maximize the area so that we can prove--o.k. we have the given base--we'll set our base equal to something.

83. K: Yeah, mumbling, P, or something--I don't know.



84. D: Then the other two sides have to add up to P.

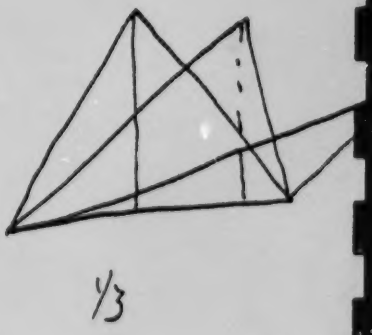
85. K: We--how about we say--let's start with an equilateral, just for the hell of it--see what happens. You get 1/3P, 1/3P and 1/3P. And this is 1/9 - 1/36 which is the height--

86. D: Now the thing we want to do is say--o.k. if we shorten this side at all and then what's going to happen to the height--if we leave this the same.

87. K: We can't shorten it.

88. D: And we shorten this side--sure we can--

89. K: Well--



90. D: We can have a--this equal to 1/3 and then a--this equal to--well you're going to have--I mean--

91. K: Aha.

92. D: This is going to get longer like that. Now we can see from this that all that is going to happen is that the base is going to get shorter so we know from that as far as leaving the base constant goes if we move--if we shorten this side then it is going to--somehow the point's going to go down in either direction.

93. K: Semicircle.

94. D: Right. That proves that we have to have an equilateral.

95. K: No, it proves an isosceles.

96. D: No, isosceles, I mean. All right from that if we set--we know that those two have to be equal so if we set this base equal to anything--it doesn't have to be 1/3P--we can also show that if this goes down--the area is going to get smaller, so the constant base then the height is going to get shorter and shorter and is getting smaller and smaller actually.

97. K: O.k., o.k.

98. D: In this case if it goes down to this side, we're going to have again a smaller angle here, shorter base here--and [noise].

99. K: So we get--so we know it is an equilateral--well prove it.

100. D: I don't know that's not a rigorous proof, but it is a proof--good enough for me.

101. K: Proves that an equilateral has the largest area.

102. D: Oh, we're talking about the largest area.

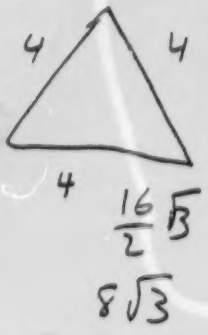
103. K: Yeah.

104. D: Oh, we just did.

105. K: We have to prove it has fixed number P--perimeter.

106. D: Well we already--we assumed that we have a fixed P, all right? I mean this is a proof as far as I.

107. K: Well, we've shown that an equilateral has the largest area. We haven't shown that if you have a certain set perimeter, let's say a right triangle, with a perimeter which is the same--we will not have a larger area.

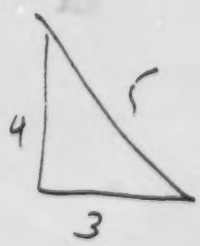


108. D: No, but we have because we have shown with the set perimeter--o.k. we know that--

109. K: Well what if we have 3, 4, 5 with an equilateral being 4, 4, 4--

110. D: 3, 4, 5 is what? Mumbling.

111. K: 12. So this area will be 6 and this area will be side squared 16. --o.k. that will have the largest area.



112. D: What's that 1.7?

113. K: Yeah, 8 is still greater than 6 and that's greater than 1.

114. D: Oh, yeah, that's right. Yeah, but the thing is if we have a fixed dimension, we already showed that, o.k. what is going to happen is as this

side gets longer--say we use 4 as a base here, so then what's going to happen--well say we use 3 as a base, just so we won't have an equilateral when we are done--what's going to happen as 4 gets longer and 5 gets shorter--it's going to go upwards. The optimum area--the maximum area is going to be right there. Because you've got--

115. K: Right.

116. D: This angle and that height. If you make this angle any less--maybe let me draw a picture--

117. K: I can understand that--this will give us largest area, but how can we prove this bottom is one-quarter--1/3 the area of the perimeter?

118. D: Well, remember all the problems we've done where we say--o.k. let me just start from here once more--so that we have 3, 4, 5--is that what you have--because that's going to be 5. Wasn't a very good 3, 4, 5 anyway. So you start out with 3, 4, 5--all right, we pick the 3 has the base, right?

119. K: Aha.

120. D: All right, it's 5--mumbling--if we have 3 as the base--and this is a little bit off an isosceles, but if we draw an isosceles as 3 as the base--o.k. we've got a right angle--that's got to be the maximum--mumbling--(height?) because if it goes any--

121. K: Right.

122. D: Over this way, it is going to go down.

123. K: O.k.

124. D: All right, so remember the argument we've used--well if we--

125. K: Yeah, I can show that, but what you're not showing is--what you're not proving is that--

126. D: That it has to be an equilateral?

127. K: Right. But you're not showing that this side is 1/3 the perimeter.

128. D: Right. I'm showing--first of all it has to be an isosceles. Right.

129. K: Right.

130. D: It has to be an isosceles--that means that we've got these three sides and those two are equal--right?

131. K: Umma.

132. D: Right--so now I pick this side as my base--I already picked--if that side is my base then the maximum area would have to have an isosceles--so I turn around--this side is my--

133. K: That I understand as proof, but you're not showing me that this is 1/3 the perimeter--mumbling.

134. D: If we have an isosceles triangle--if we have an equilateral triangle--then each side has to be 1/3 the perimeter--that's the whole thing about an equilateral triangle.

135. K: I know--o.k.

136. D: First we know it must be an isosceles, right?

137. K: Umma.

138. D: O.k.

139. K: I understand this.

140. D: If it is an isosceles, it must be an equilateral, right?

141. K: All right.

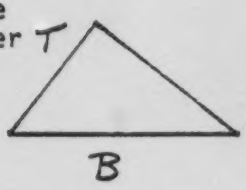
142. D: And if it must be an equilateral--all three sides must be equal and if the perimeter is P, all three sides must be 1/3P.

143. K: O.k. I've got it.

Appendix 4

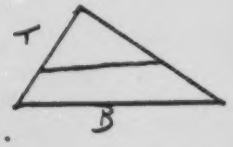
Protocol 4

1. (Reads problem) You are given a fixed triangle T with base B. Show that it is always possible to construct, with ruler and compass, a straight line parallel to B such that that line divides T into two parts of equal area. Can you similarly divide T into five parts of equal area?



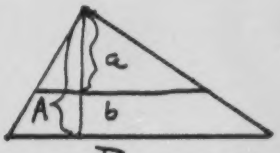
2. Hmmm. I don't know exactly where to start.

3. Well I know that the...there's a line in there somewhere. Let me see how I'm going to do it. It's just a fixed triangle. Got to be some information missing here. T with base B. Got to do a parallel line. Hmmm.



4. It said the line divides T into two parts of equal area. Hmmm. Well, I guess I have to get a handle on area measurement here. So, what I want to do...is to construct a line... such that I know the relationship of the base...of the little triangle to the big one.

5. Now let's see. Let's assume I just draw a parallel line that looks about right, and it will have base little b.



6. Now, those triangles are <u>similar</u>.

7. Yeah, all right then I have an altitude for the big triangle and an altitude for the little triangle so I have little a is to big A as little b is to big B. So what I want to have happen is 1/2 ba=1/2AB-1/2ba. Isn't that what I want?

$\frac{a}{A} = \frac{b}{B}$

8. Right! In other words I want ab=1/2AB. Which is 1/4 of A times...mumbles(confused)...One over the square root of two times A times one over root two times B.

$\frac{1}{2}b \cdot a = \frac{1}{2}A \cdot B - \frac{1}{2}ab$

$ab = \frac{1}{2}AB - \left(\frac{1}{\sqrt{2}}A\right)\left(\frac{1}{\sqrt{2}}B\right)$

9. So if I can construct the square root of two, which I can! Then I should be able to draw this line...through a point which intersects an altitude dropped from the vertex. That's little $a=A/\sqrt{2}$, or $A=a\sqrt{2}$, either way.

10. And I think I can do things like that because if I remember I take these 45^0 angle things and I go $1,1,\sqrt{2}$.

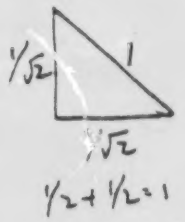
11. And if I want to have a times root 2...then I do that... mmm...wait a minute...I can try and figure out how to

$\frac{1}{\sqrt{2}}A = a$

$A = \sqrt{2}a$

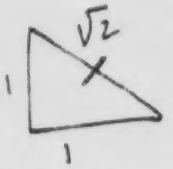
construct $1/\sqrt{2}$.

12. O.k. So I just got to remember how to make this construction. So I want to draw this line through this point and I want this animal to be...$1/\sqrt{2}$ times A. I know what A is, that's given. So all I got to do is figure out how to multiply $1/\sqrt{2}$ times it.



13. Let me think of it. Ah huh! Ah huh! Ah huh! $1/\sqrt{2}$...let me see here...ummm...that's 1/2 plus 1/2 is one...

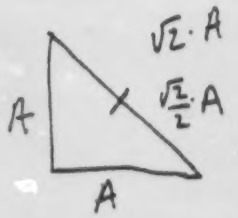
14. So of course if I have a hypotenuse of one...



15. Wait a minute: $1/\sqrt{2} \cdot \sqrt{2}/\sqrt{2} = \sqrt{2}/2$...that's dumb!

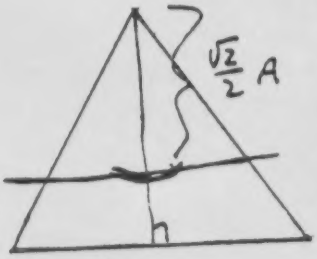
16. Yeah, so I construct $\sqrt{2}$ from a 45, 45, 90. O.k. so that's an easier way. Right?

17. I bisect it. That gives me root 2 over 2. I multiply it by A...now how did I used to do that?



18. Oh heavens! How did we used to multiply times A. That... the best way to do that is to construct A...A...then we get root 2 times A, and then we just bisect that and we get $A\sqrt{2}/2$. O.k.

19. That will be...what!...mmm...that will be the length...now I drop a perpendicular from here to here. O.k....and that will be...ta, ta...little a.

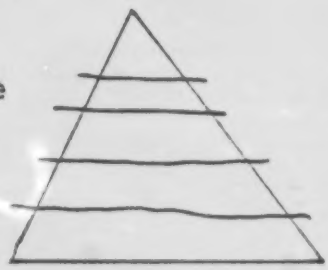


20. So that I will mark off little a as being $A\sqrt{2}/2$. O.k. and automatically when I draw a line through that point...I'd better get $\sqrt{2}/2$ times big B. O.k.

21. And when I multiply those guys together I get 2 over 4 times A times B. So I get half the area...what?...yeah...times 1/2...so I get exactly 1/2 the area in the top triangle so I better have half the area left in the bottom one. O.k.

22. O.k., now can I do it with 5 parts?

23. Assuming 4 lines.

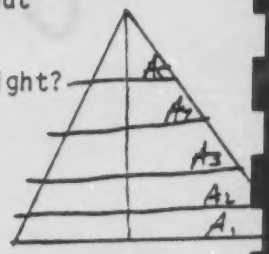


24. Now this is going to be interesting since these lines are going to have to be graduated...that...

25. I think, I think, that rather then get a whole lot of triangles here, I think the idea, the essential question is can I slice off...1/5 of the area...mmm...

26. Now wait a minute! This is interesting. Let's get a...how about four lines instead of...

27. I want these to be...all equal areas...right? A_1, A_2, A_3, A_4, A_5 right?



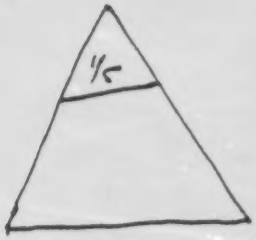
28. Sneak! I can...I can do it for a power of 2...that's easy because I can just do what I did at the beginning and keep slicing it in half all the time.

29. Now can I use that kind of induction thought.

30. I want that to be 2/5. And that to be 3/5.

31. So let's make a little simpler one here.

32. If you could do that then you can construct the square root of five. But I can construct the square root of 5 to one...square root of 5, right?

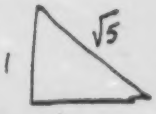


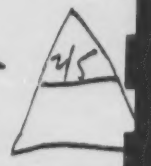
33. So I can construct...o.k. So that certainly isn't going to do it. No contradiction...

$a \cdot b = \frac{1}{5} A \cdot B = \frac{1}{\sqrt{5}} A \cdot \frac{1}{\sqrt{5}} B$

34. Now, I do want to see, therefore, what I have here.

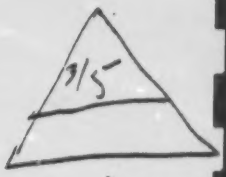
35. I'm essentially saying is it possible for me to construct it in such a way that that is 1, 2, 3, 4, 5, 1/5 the area...o.k.





36. So little a times little b has got to equal 1/5 times A times B. So I can certainly chop the top piece off and have it be 1/5 of the area. Right? Right?

37. Now, from the first part of the problem...I know the ratio of the next base to draw...because it is going to be root 2 times this base. So I can certainly chop off the top two fifths.



38. Now, from the first part of the problem I know the ratio of the top...uh, o.k. now this is 2/5 here, so top 4/5... o.k....all right...so all I got to be able to do is chop off the top 3/5 and I'm done...

39. It would seem now that it seems more possible...let's see...

40. We want to make a base here such that little a times little b is equal to...the area of this thing is going to be 3/5... 3/5AB...in areas, right!...and that means little a times little b is $\sqrt{3}/\sqrt{5}$A times $\sqrt{3}/\sqrt{5}$B. O.k. then can I construct the square root of $\sqrt{3/5}$. If so then this can be done in one shot.

$ab = \frac{3}{5} A \cdot B =$

$\left(\frac{\sqrt{3}}{\sqrt{5}} A\right)\left(\frac{\sqrt{3}}{\sqrt{5}} B\right)$

68

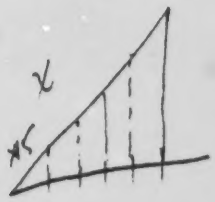
41. Well let's see. Can I construct $\sqrt{3/5}$. That's the question. $\sqrt{3}/\sqrt{5} \cdot \sqrt{5}/\sqrt{5} = \sqrt{15}/5$.

$\frac{\sqrt{3}}{\sqrt{5}} \cdot \frac{\sqrt{5}}{\sqrt{5}} = \frac{\sqrt{15}}{5}$

42. Root 15, root 15. Wait a minute! Root 15 over 5. Is the square root of 15 constructable? Root 15 is...

$\sqrt{16-1}$

43. It is the square root of 15-1. But I don't like that. It doesn't seem the way to go.



44. $16^2 - 1^2$ equals... (expletive deleted)

45. Somehow it rests on that.

46. (expletive) If I can do the square root of 15. Can I divide things and get this?

47. Yeah, there is a trick! What you do is you lay off 5 things. 1, 2, 3, 4, 5. And then you draw these parallel lines by dividing them into fifths. So I can divide things into fifths so that's not a problem.

48. So it's just constructing the square root of 15 then I can answer the whole problem.

49. I got to think of a better way to construct the square root of 15 then what I'm thinking of...or I got to think of a way to convince myself that I can't...umm...x^2-15.

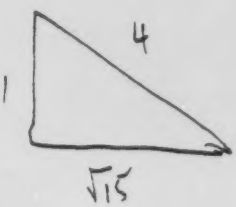
50. Trying to remember my algebra to knock this off with a sledge-hammer.

51. It's been so many years since I taught that course. It's 5 years. I can't remember it.

52. Wait a minute! Wait a minute!

53. I seem to have in my head somewhere a memory about quadratic extension.

54. Try it differently here. mmm...



55. So if I take a line of length one and a line of length... And I erect a perpendicular and swing a 16 (transcriber's note: for mathematical clarity he really means 4 instead of 16) here...then I'll get the square root of 15 here, won't I?

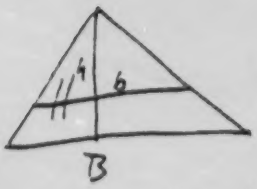
56. I'll have to, so that I can construct the square root of 15 times anything because I'll just multiply this by A and this by A and this gets multiplied by A divided by 5 using that trick. Which means that I should be able to construct this length and if I can construct this length then I can mark it off on here and I can draw this line and so I will answer the question as YES!!

Appendix 5

Protocol 5

1. (Reads problem) Same as Protocol 4.

2. The first thought is that the two triangles for the first question will be similar.



$bh = \frac{1}{2}BH = \frac{1}{\sqrt{2}}B \cdot \frac{1}{\sqrt{2}}H$

3. And since we'll want the area to be one half. And area is related to the product of the altitude and the base we want the area of the smaller triangle to be one half.

4. And corresponding parts of similar triangles are proportional. We want the ratio of proportionality between the altitudes and the bases both to be $1/\sqrt{2}$.

5. So I will draw a diagram...and I'm drawing that parallel and checking that algebra.

6. I hope you can hear the pencil moving because that's what's happening at this point.

7. And now I'm writing a bunch of letters on my diagram and multiplying them together...leaving the one half out, of course...and I want that to be one half of that.

8. So, that certainly seems like a reasonable solution. So all I have to be able to do is construct $\sqrt{2}$. And I can do that with a 45 right triangle, and then given a certain length, namely the altitude, to the base B, which I can find by dropping a perpendicular. I want to construct a length which is $1/\sqrt{2}$ times that, and I can do that with the ordinary construction for multiplication of numbers.

9. So, I can do the problem.

   I: You can do all the constructions?

10. Yeah, I do them in the winter term. This line, this line, here's one, you want to multiply p times q, you draw these parallels and it's pq.

    (The solution of part 2 is omitted)

On the Analysis of Two-Person Problem Solving Protocols

Alan H. Schoenfeld
Graduate School of Education and
Department of Mathematics
The University of Rochester
Rochester, NY 14627

SE038834

Paper given at the 1982 annual American Educational Research Association meetings, New York, March 1982.

The research described in this paper was supported by a National Science Foundation grant, SED 79-19049.

To appear in the Journal of Mathematical Behavior.

1. Overview

This paper is one of a pair which, together, try to sketch out some of the issues that should be taken into account when one uses certain "verbal methods" (clinical methodologies or protocol analysis) for research into human problem solving processes. This paper is primarily a case study in one methodology, in which two students are videotaped as they work together to solve mathematical problems "out loud." The focus will be on the advantages, and disadvantages, of this particular methodology -- or more properly, on those aspects of cognitive processes that this methodology will illuminate and those which it will obscure. The context for this discussion is treated at some length in the companion paper, "Beyond the purely cognitive: Meta-cognition and social cognition as driving forces in intellectual performance." A brief discussion of that context is given in section 3.

2. Background

In recent years there has been a resurgence of the use of verbal data for research into the nature of human cognitive processes. Such research takes as its data the verbal reports produced by individuals or groups of subjects in a variety of circumstances: through retrospection or introspection, in structured or unstructured "clinical" interviews, in "speak aloud" problem solving sessions, with or without experimenter intervention. Through the period of the Gestaltists' major influence, the analysis of verbal reports or introspections was considered methodologically sound, if not the primary source of information regarding complex human cognitions. However, verbal methodologies fell out of favor with the advent of behaviorism and the rise of "scientific" methodologies for the investigation of cognitive phenomena.

The mental constructs posited by the Gestaltists were unneccessary for (or more accurately, antithetical to) the theoretical foundations of the behaviorists (see, for example, Skinner [1974]). In addition the products of introspection were not replicable or verifiable. Perhaps more importantly, they were not falsifiable. Thus they could not, it seemed, serve as the foundation for a cumulative scientific effort. In consequence the methodologies that gave rise to such unscientific results were supplanted by more "rigorous" methodologies that promised to yield "good science." Verbal methods were déclassé through the 1960's and the 1970's.

For a number of reasons, perspectives on verbal methods have changed in recent years. Perhaps the major cause of the change is the "legitimization" of protocol analysis as a consequence of its role as a major research tool in artificial intelligence. Such research (see, for example, Newell and Simon [1972]) demonstrated that one can design successful problem solving programs for computers, based on principles abstracted from the analysis of human problem solving protocols. These computer programs offered, for the first time, incontrovertible empirical "proof" of the efficacy of certain strategies*, and gave credibility to the methodologies that uncovered them. Another major cause was the impact of Piaget's genetic epistemology in general and, in mathematics education, the impact of Krutetskii's work (Krutetskii, 1976). Piaget's work made it clear that careful clinical investigations could give rise to replicable results, to falsifiable hypotheses, and to predictions that could be tested experimentally. In short, clinical investigations could indeed lay the foundations for good science. Krutetskii's

*Technically, they offered proof that the machine implementation of such strategies is possible, not that humans actually use the strategies.

work was not "science" in the unbiased, objective sense that we take it. However, it dealt with issues from a perspective that appeared to provide more direct explanations of students' "real" mathematical behavior than the work coming from many "scientific" studies. This caused much interest in his methodologies (and Soviet "teaching experiments" in general.)

Indeed, one cause of the resurgence of verbal methods is the increased sophistication of the research community and its more balanced perspectives on the methodologies that supplanted them. The limitations of the statistical methodologies began to emerge as it became clear, from a lack of clear-cut results in the empirical literature, that there are often (for example) many more variables in "treatment X vs. treatment Y" comparisons than are being controlled for in supposedly "tight" experimental designs. It became clear as well that the difficulties in extrapolating results from well designed laboratory studies to more complex cognitive phenomena, and to more complex environments, had been seriously underestimated. Calls were made (e.g. Kilpatrick, 1975) for the use of clinical investigations to determine, in exploratory fashion, the spectrum of important "mathematical abilities." More recently, the cognitive community has begun to recognize the importance of "other than purely cognitive" influences on what were once taken as "purely cognitive" actions. Thus the role of metacognitions and social cognitions (belief systems, etc.) as driving forces in human intellectual performance is coming to receive more attention (see, e.g., Brown, 1978; D'Andrade, 1981; Lawson, 1980). A range of exploratory methodologies, often verbal, has been developed to deal with such questions. Hence for many different reasons, verbal (clinical or protocol) methods are used with increasing frequency as research tools. Yet,"while increasingly popular, protocol

methods have not yet received thorough methodological analysis. Little is known concerning their fundamental natures, the rationales underlying their use, and their reliability" (Ginsburg, Kossan, Schwartz, and Swanson, in press). Such analyses are beginning to emerge, the Ginsburg et al paper being one of them. Also, Psychological Review has published two recent analyses of the effects of speaking aloud methods. Nisbett and Wilson's [1977] title, "Telling more than we know: Verbal reports on mental processes" suggests its conclusions. Ericsson and Simon [1980] conclude that certain kinds of "talking aloud" instructions -- those that ask for verbalization as one solves a problem, without calling for explanations (elaborations or retrospections) of what one is doing -- do not seem to affect people's performance while solving problems. This paper and its companion will suggest that that conclusion needs to be further qualified. Some of the relevant issues and variables are characterized next.

3. Context

Issues regarding the validity and generality of verbal methods are singularly complex and subtle. Any particular framework for gathering and analyzing verbal data will illuminate certain aspects of cognitive processes and obscure others.* Perhaps more importantly, it may appear to illuminate many behaviors that have, in actuality, been distorted in a number of subtle ways. Each methodology is a lens (or filter, if you will) through which intellectual performance is being viewed. Thus the selection of any particular methodology for investigation may well determine what the experimenter does

*To be accurate, I should talk of "social-cognitive" processes, in the sense that the "cognitions" being studied take place in a social context, which may well determine what the experimenter sees. The discussion below will clarify this point.

or does not see. In turn, this may affect the theoretical constructs that are derived from these observations. Since there is great potential for distortion in this arena, the experimenter wishing a sense of the "whole cognitive picture" should consider using a range of complementary (verbal and other) methodologies, and must be extremely cautious in interpreting the results obtained from a body of methodologically similar studies.

A wide range of variables affect the kinds of information that emerge from verbal methodologies. Some of them are sketched briefly here.

a. The number of persons being taped.

Radically different types of behavior emerge in single-person, two-person, and small group (say three to five people) protocols. The prevailing assumption is that single-person protocols give rise to the "purest" cognitions, uncontaminated by social concerns. However, the task environment itself imposes certain constraints upon the subject(s),** and the discomfiting effects of a task environment may be strongest when a person is solving a problem alone, rather than with the (intellectual and social) support of a peer. Certain behaviors become more prominent, and easier to observe, with more than one subject (e.g. decisionmaking). However, observing other aspects of behavior is made more difficult. One dominant member of a group can skew discussions to the point where they reflect only that person's ideas; solutions may proceed in parallel, or with (positive or negative) reinforcement from the interactions. The more people involved, the more obvious the social dynamics. There are no value judgments attached to these characterizations -- each serves its purpose, and one should simply choose

*These "environmental" constraints lessen with the maturity and training of the subjects. However (see below) college seniors still feel them strongly.

the one(s) suited to the ends that one has in mind. If one is interested in making artificial intelligence models of competent problem solving performance, [e.g. Newell and Simon, 1972] then the most appropriate methodology may well be to perform the detailed analyses of single-person protocols. If one wishes to elucidate certain kinds of decisionmaking behavior, [e.g. Schoenfeld, in press] two-person protocols may be appropriate. If one wishes to make statements about students' "real, social" cognitive behavior [e.g. Lesh's "applied problem solving project" and Noddings' analyses of group interactions] then larger groups are appropriate.

b. The degree of intervention

Verbal methods include a continuum of experimenter obtrusiveness that ranges from near invisibility (covert or non-interventionist observations of people in natural settings) to positions of central importance (experimenters inducing "cognitive dissonance" in clinical interview settings). Each serves certain ends in particular situations. If, for example, an experimenter is interested in determining the "Van Hiele level" of a student on geometry tasks (or the Piagetian level of a subject on a particular task), and exploring corollary behavior on other tasks, then a large degree of intervention is almost mandatory. If, however, the experimenter wishes to see how a student copes with difficult problems (what the student pursues, whether the student goes off on "wild goose chases," etc.), then intervention may be inappropriate. Indeed, asking the student "why did you do X?" may have a dramatic effect on the student's behavior. Up to that point, the student may not have considered the question. There is, first, the chance that the answer to the question is "manufactured." Second and equally important, the student is now aware that the experimenter is interested in

how such choices are made. The student may begin to reflect on on those choices while working on the given task, and behave from that point on in a manner very different than he or she would otherwise have behaved.

c. The nature and degrees of freedom in instructions and intervention

The kind of instructions subjects are given has a strong effect on what they produce. For example, asking the subject to reflect upon his or her problem solving processes does have an effect on performance [Ericsson and Simon, 1980]. Yet such reflection may point to behaviors that might otherwise be unseen. In clinical interviews there are tradeoffs between standardization on the one hand and experimenter freedom on the other; one has a certain degree of reliability in the first case, and the potential for probing interesting behaviors in the second.

d. The nature of the environment and how comfortable the subject feels in it

To put it simply, students who feel uncomfortable in a particular environment may uniformly exhibit pathological behavior. That the behavior is pathological may not at all be apparant; that may only appear when the experimental conditions are altered. Further, putting students "at ease" may be completely insufficient. The very fact that one is being taped may be enough to induce atypical behavior (see below). Subjects may avoid dealing with the task in any substantive way, in order to avoid feelings of inadequacy when they (as they see it, inevitably) fail at it. They may create certain kinds of behavior, to make it seem as if they know what they are doing. They may select their behavior to tailor it to (what they believe are) the experimenter's wishes. (In the later category, I have tapes in which students say "We could solve it like this, but obviously he doesn't want that.") Or, students may deal with a problem in unusual ways simply because they are in an obviously artificial setting. (A student in one of

Dick Lesh's videotapes, working on a "real world" problem, misread some given information and assumed that he could earn nearly $150 for mowing one person's lawn once. When he was questioned later, he was asked if that seemed like a reasonable figure. It did not. In fact, the student mowed lawns for extra money and knew the figure was unreasonable. But, "it was a hypothetical question, wasn't it?")

e. Task variables

The range of these is tremendous. Does one provide children with manipulatives, for example? How does this affect performance? For a general discussion of task variables in mathematical problem solving, see Goldin and McClintock [1979].

This brief discussion serves to indicate some of the variables that affect the collection and interpretation of verbal data. It is a bare introduction to an area that needs much greater investigation, but it may serve to set the stage for the following discussions.

4. Executive decisions in problem solving: the issue and methodology

As section 3 indicates, one's choice of methodology should be guided by the goals one has for research. The "problem" I set out to investigate, initially, was to explore some of the reasons for students' lack of success in solving "non-routine" problems at the college level. In addition, I wished to examine students' performance before and after a course in mathematical problem solving, in order to determine some of the effects of instruction. Previous work had provided some tools for the investigation, and some ideas as to what mechanisms might contribute to success (or more accurately, to failure). The general arena was an investigation of Pólya-type heuristics and their contributions to problem solving performance.

Earlier studies had indicated that students could learn to use individual heuristics with some degree of competency [Schoenfeld, 1979] and a battery of tests had been used to examine fluency and competency in the implementation of those heuristics. Thus the intention here was not to investigate such competencies. (If it were, some form of detailed clinical probing would undoubtedly have been an appropriate methodology.) Rather, the intention was to investigate a consistent "difficulty" with regard to heuristics. The literature indicated that, while students did seem able to master individual problem solving strategies, the overall effects on their problem solving performance was not nearly as large as was expected or hoped: the problem solving whole was, somehow, less than the sum of its heuristic parts [Wilson, 1967; Smith, 1973; Lucas and Loomer (in Harvey and Romberg), 1980; Goldberg, 1975]. The questions chosen for investigation were: what will students choose to examine in a problem solving situation (and why)? How will they "follow up" on those choices (pursue them, abandon them etc.)? and What is the effect of such "strategic" or "executive" behavior on their problem solving performance? Observe that these questions can be asked about problems that may or may not be amenable to particular heuristic strategies for solution, solved by students who may or may not have the heuristics at their disposal. This was an exploratory study, in that the data (videotapes) were to serve as a source of hypotheses rather than as a test of them. Some of the choices among the variables given above, and the rationales for them, were as follows.

a. The number of persons being taped

For a variety of reasons, two-person protocols provide the richest data for the purpose described above. First, I have found that single-person

protocols (from students, not faculty) tend to generate unnatural behavior in subtle ways. Protocol 1 (appendix 1) was generated by a single student, a senior mathematics major. It is typical of single-person protocols for this problem ("How many cells are there in an average adult human body?") in that much time and effort is spent approximating parts of the body by geometric solids and computing the volume of those solids. In roughly two dozen two-person protocols, not one pair of students has done the same.* This behavior was induced by the setting: the students felt the need to "produce something mathematical" for the researcher. Many-person protocols ease the pressure on the subjects, for the burden of uncomfortableness is shared among the students.

A second reason for not using single-person protocols in these circumstances is the wish to elucidate the nature of the students' strategic decision-making as they work on the problems. For reasons given below, the sessions had little or no experimenter intervention and the subjects were not instructed to explain what they did as they solved problems. In single person "speak aloud" protocols, what appears is often the "trace" of a solution: one sees the results of decisions but gets little insight into how the decisions were made, what options were considered and rejected, etc. When students work together as a team, discussions between them regarding what they should do next often bring those decisions and the reasons for them "out in the open." (A typical dialogue is "Let's do X." "Why? I don't see what good it'll do." "Look...")

*I collected the single-person protocols first, and had begun to construct various (cognitive) explanations for this poor strategic behavior. Only later, when there were two-person protocols for comparison, did it become apparent that the extensive body-volume computations were caused by the social environment.

These reasons suggested the use of many-person ($n \geq 2$) protocols. There are trade-offs with regard to group size. Larger groups provide more "ideas" to manage, and decision making can be more interesting in these circumstances. Also, social dynamics of groups of 4 or 5 may better ameliorate the uncomfortableness of the experimental environment. Two reasons suggested $n = 2$ as the most suitable choice. First the decisions one faces when "managing" the ideas generated by a group of people may be very different from the decisions one faces when considering the ideas one or two people have generated. For example, one or two people working alone might go off on a "wild goose chase" and squander their problem solving resources that way. In a larger group, the likelihood of someone saying "why?" to the proposed direction is greater, and the resultant behavior different. Also, one or two students may only generate one or two plausible alternatives; a "committee" may generate more. The "perceived solution space" is different, and the resulting behaviors may not reflect those of individuals working alone or in pairs. Second, the focus of this investigation was largely cognitive. With larger groups the degree of social interactions increases, making it more difficult to tease out the "purely cognitive" aspects of students' behavior. These social factors are still (all too) present in 1- and 2-person protocols, however. We have seen how one person "engages in mathematical behavior" to dissipate the pressure of the task environment. In similar circumstances a pair of students may "defuse" the environment by engaging in small talk around the problem. By refusing to take it seriously, they can justify (what is from their perspective) the inevitable failure by telling themselves that they "never really tried." And of course, two-person social dynamics can be quite strong. In all cases, one must take care that the behavior

labeled as "cognitive" is indeed so.

b. The degree of intervention

It is important to keep interventions to an absolute minimum in this kind of study. The idea was to determine the presence (or absence) of certain kinds of "monitoring" and decision making in students' problem solving, and to trace the effects of their presence (or absence). These effects can only be seen if a solution is allowed to run its course. For example, a student may have a "hunch" or some "intuition" about a plausible solution, and begin to work in that direction. From the experimenter's point of view, it may be clear that this is a "wild goose chase," and it may be tempting to find out what prompted the student to pursue it. However, such an intervention precludes the opportunity to observe the effects of such a wild goose chase. After three minutes the student might come to see that it is fruitless, and go on to do something else. Or, the student might never reconsider, and spend the allotted time involved in irrelevancies. In fact, the latter type of behavior occurs all too frequently. In a large number of tapes, students engaged in an essentially irrelevant computation for nearly twenty minutes (the length of the taping sessions). After they ran out of time they were asked what they would do with the result of the computation if it were given to them...and they were unable to say [Schoenfeld, in press]. I now believe that this lack of monitoring is quite typical (though not always this extreme, obviously) of student behavior, and is one of the major contributing factors in students' problem solving failures. This could only be seen, and verified, by letting the solutions proceed unimpeded.

More importantly, in this particular kind of study, experimenter intervention may have a radical effect on the subject's performance and on the

data that is produced. Recall that one purpose of these experiments was to determine the degree to which students reflect on, and oversee, the way that a solution evolves. Suppose, for example, that a student in the midst of a solution is asked to justify a "wild goose chase" or any other strategic decision. Up to that point the student may not have thought about such issues, or dealt with them casually. After the intervention, he or she knows that the experimenter is interested in such questions. In consequence, the student may begin to manufacture such justifications (to be ready for the next intervention). In doing so, that person's behavior may be completely distorted. There is now a training effect, and all data must be interpreted accordingly.

Again, the preceding comments should be interpreted in the context of the goals for the study, which was exploratory. One of its purposes was to explore students' monitoring and executive behaviors, and document the role that they play in students' problem solving performance. Once that information has been gathered satisfactorily, a shift in methodology may be appropriate. The fianl section will mention a revised methodology I am now pursuing.

c. The nature of instructions and interventions

As the reports in Psychological Review [Nisbett and Wilson, 1977; Ericsson and Simon, 1980] indicate, asking students to talk about their problem solving processes during the process of a solution does have an effect on those processes. For that reason students were instructed not to "explain" for the microphone. Rather, they were to work together as a team; what I wanted to hear would emerge from their discussions. Interventions were limited to comments like "I haven't heard you say much in the past three minutes. Are you still working on the problem?" and to specific

responses to questions the students asked.

d. The environment

The setting was obviously artificial (solving non-routine problems for a mathematics professor can hardly be considered natural behavior) and, despite all efforts to the contrary, somewhat stressful. This point should be emphasized. Protocol 1 (appendix 1) was produced by a senior mathematics major who was on a first-name basis with the experimenter. The student was familiar with the entire process (he had done some taping himself as part of a senior thesis). Yet the unfamiliar problem induced great stress, with a resultant effect on the protocol. Similarly, great efforts were taken in all the two-person protocols to put the students at their ease. They were assured that the research was non-judgmental, shown that the videotape machine focused on the pages they were writing and not on their faces, etc. Even so, it is not at all safe to assume that the students would deal with the same problems in anything like the same manner if they worked on them, for example, in their own rooms without a recording device present. The more awkward the situation -- the more obtrusive the recording equipment, the more "unusual" the problem, etc. -- the more likely the "verbal data" is to be affected.

e. Task variables

The subjects were college students, and treated as such. Problems were worked with paper and pencil only, save for "straightedge and compass" geometry constructions, where they were given the tools for the constructions.

5. A framework for examining the protocols

An extensive description of the framework described below, and of the results obtained with it, is given in my article "Episodes and Executive

Decisions in Mathematical Problem Solving" [in press]. In brief, the idea was to create a macroscopic framework that captured the essential elements in the problem solution.* There is one significant difficulty in implementing this idea, a difficulty that has been the downfall of most extant protocol coding schemes: the most important event in a problem solving session may be one that is conspicuous by its absence! For example, appendix 1 of the "Episodes" paper gives the protocol of a tape that had a 20-minute long "wild goose chase" in which the students tried to calculate the area of a geometric figure. At the end of the tape they were asked how they would use the result if they had it, and they could not say. Had they asked themselves, at the moment they set out to do the calculation, what value it would have, they might have avoided wasting their time. But they did not, and the solution was doomed from that point on. Now observe that, as one might expect, conventional coding schemes record overt behaviors in a problem protocol. While this seems to be perfectly natural, the result is that such frameworks bypass the critical element in the protocol described above: the absence of evaluation at a "make or break" point in a solution. Such systems will not point to the reason that the attempt failed. The general idea is to discuss the impact of the (presence or absence of) assessments and consequent decision making of the solution as a whole.

The idea behind the generation of the system is straightforward. Potential "make or break" points in a solution occur whenever the direction of a solution changes radically (when one approach is abandoned for another), or when new

*A qualitative "test" for capturing the essential elements in a protocol is the following. After being given the analysis (coding) of the protocol, are there "surprises" when one sees the tape for the first time? This particular framework seems to pass the test. The other systems with which I am familiar fail it miserably: it is nearly impossible to get a "feel" for what happened from the string of coding symbols.

information arises that might call for such a radical change. The system is designed to identify those points, to characterize the behavior of the students at those points, and to describe the effect of that behavior on the solution.

A protocol is parsed into major "episodes." An episode represents a body of consistent behavior on the part of the problem solver(s). Episodes have one of six labels attached to them: reading, analysis, exploration, planning/implementation, verification, and transition. Once a protocol has been parsed into episodes, one category of "make or break" points becomes obvious: any transition point between episodes is a potential assessment/decision point. Other "executive" decisions should be made at "new information" points.

Appendix 2 provides the full analysis of a protocol, which is given in appendix 3. This analysis provides an example of how the system works. Most of the commentary is self-explanatory.*

6. Discussion

The framework discussed above has proven itself reliable and, I believe, reasonable informative. It seems to capture much of the "essence" of a problem solving session, without getting lost in details. The macroscopic approach allows one to get a sense of apparent causes of success or failure in a

*Letters preceding comments refer to specific parts of the framework. For example, one asks three questions about any reading episode:

a. Have all of the conditions of the problem been noted? Explicitly or implicitly?
b. Has the goal state been correctly noted? (Again, explicitly or implicitly?)
c. Is there an assessment of the current state of the problem solver's knowledge relative to the problem solving task?

In virtually all cases, the questions that lie behind the comments in appendix 2 are clear. They are omitted to save space.

problem solution, and points out the importance both of monitoring solutions and of "executive" decisions in them. The framework is straightforward to implement (three students, in concert, do most of the coding for me) and reliable (their codings and mine have an intercoder reliability exceeding 85%). The framework is also generalizable: it is not domain-specific, and can be adapted easily to study problem solving behavior in other disciplines. However, some caveats are in order.

First of all, this particular methodology offers only one perspective on the problem solving process. It should be coupled with a variety of others (paper-and-pencil tests, clinical interviews to determine mathematical abilities, different protocol methods and different levels of analysis, etc.) in order to provide a reasonably comprehensive picture of problem solving behavior.

Second, there are any number of dangers inherent in the gathering of protocols. A few of these (for example, pathological behavior induced by uncomfortableness, or bad social dynamics) were mentioned above. It is nearly impossible to control for these, or even to be aware of them in any particular protocol. Thus one must exercise extreme caution in providing "purely cognitive" explanations for behavior.

Third, this was an exploratory methodology and has certain limitations. It was non-interventionist, for example, in order to make the case that assessments and managerial decisions play a critical role in determining the success of problem solving attempts. Once that point has been granted, one may well wish to explore "executive" behavior in more detail. I am now trying a variant of the preceding methodology, as follows. A student is first videotaped in the fashion described above. Then the student watches

the videotape and critiques it, explaining the reasons for his or her behavior. These explanations are probed in clinical fashion. This "mixed" methodology will, I hope, allow for a more subtle elucidation of problem solving processes.

Summary

This paper has discussed some of the subtleties involved in the use of verbal methodologies. It has examined in some detail the aims and rationales of a particular methodology, two-person speak-aloud protocols without experimenter intervention, and discussed a framework for analyzing such protocols. Verbal methodologies, if used with care, can help to shed much light on cognitive processes. It is hoped that this is a step in that direction.

References

Brown, A. Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in Instructional Psychology. (Vol. 1) New Jersey: Lawrence Earlbaum, 1978.

D'Andrade, R.G. The cultural part of cognition. Cognitive Science, 1981, 5, 179-95.

Ericsson, K.A. and Simon, H.A. Verbal reports as data. Psychological Review, 1980, 87(3), 215-51.

Ginsburg, H., Kossan, N., Schwartz, R., and Swanson, D. Protocol methods in research on mathematical thinking. In H. Ginsburg (Ed.), The Development of Mathematical Thinking. New York: Academic Press, in press.

Goldberg, D.J. The effects of training in heuristics methods on the ability to write proofs in number theory. (Unpublished doctoral dissertation, Columbia University, 1974.) Dissertation Abstracts International, 1974, 4989B. University Microfilms 75-7,836.

Goldin, G. and McClintock, E. (Eds.) Task Variables in Mathematical Problem Solving. Columbus, OH: ERIC, 1980.

Harvey, J. and Romberg, T. Problem Solving Studies in Mathematics. Madison, WI: Wisconsin R & D Center, 1980.

Kilpatrick, J. Variables and methodologies in research on problem solving. In Hatfield, L., (Ed.) Mathematical Problem Solving. Columbus, OH: ERIC, 1978.

Krutetskii, V.A. The Psychology of Mathematical Abilities in School Children. (Translated from the Russian by Joan Teller, edited by Jeremy Kilpatrick and Isaac Wirszup.) Chicago: Chicago University Press, 1976.

Lawson, M.J. Metamemory: Making decisions about strategies. In Kirby, J.R. Biggs, J.R. (Eds.) Cognition, Development, and Instruction. New York:

Academic Press, 1980.

Newell, A. and Simon, H. Human Problem Solving. New Jersey: Prentice-Hall, 1972.

Nisbett, R.E. and Wilson, T. Telling more than we know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-60.

Schoenfeld, A.H. Episodes and executive decisions in mathematical problem solving. In Lesh, R. and Landau, M. (Eds.) Acquisition of Mathematics Concepts and Processes. New York: Academic Press, in press.

Schoenfeld, A.H. Explicit heuristic training as a variable in problem solving performance. Journal for Research in Mathematics Education, May 1979.

Schoenfeld, A.H. Measures of problem solving performance and of problem solving instruction. Journal for Research in Mathematics Education, January 1982.

Skinner, B.F. About Behaviorism. New York: Knopf, 1974.

Smith, J.P. The effect of general versus specific heuristics in mathematical problem solving tasks. (Unpublished doctoral dissertation, Columbia University, 1973.) Dissertation Abstracts International, 1973, 34, 2400A, University Microfilms, 73-26,637.

Wilson, J.W. Generality of heuristics as an instructional variable. (Unpublished doctoral dissertation, Stanford University, 1967.) Dissertation Abstracts International, 28, 1967, 2575A. University Microfilms, 67-17,526.

Appendix 1: A single-person protocol

(Reads problem): Estimate, as accurately as you can, how many cells might be in an average-sized adult human body. What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your figures?

I'll think of some approaches I might take to it.

The first one might be just to go by parts of the body that are fairly distinct and try to figure out...

My first possible approach to the problem might be to look at them as approximations to geometric shapes and try to figure out the volume of each part of the body. And then make a rough estimation of what I thought the volume of a cell was and then try to figure out how many cells fit in there.

I would say take the arm from the wrist up to shoulder and it's approximately a cylinder and it's, I don't know, about 3 or 4 inches in diameter. So you would have, it's about 2 or 1½ inches in radius, squared, times π and the volume of my arm in square inches. So I've got two arms, so I've got two of those.

And now a leg. A leg...think this might be better...there's a little more variance, so I would say a cone might be more appropriate. And the base of my leg is approximately 6 or 7 inches in diameter so you would have $3\frac{1}{2}^2 \times \pi$ and the height would be... what is my inseam size, about 32 or 34. So you've got to have a 34, and it's a cone so you've got to multiply it by 1/3.

And now the head is very, very roughly a sphere. And so you've got a sphere of... I don't know how many. I don't know, maybe on the average 6" in diameter. That may be a little small, maybe 7" in diameter. And so quick recognition of the formula was $4/3\ \pi r^3$. So I've got 4/3 of whatever my head is cubed, I've got $3\frac{1}{2}^3$, and what am I missing now?

Oh, torso...very important. Well a torso is...you could say is approximately like a cylinder except with an oval base. So I could figure out what the area may be around is, and I won't calculate this explicitly. Say my waist is about 34" and I could approximate it across here. And if I worked on it I could figure out what the geometry of it of the volume of that ellipse.

S: Well, make a ballpark estimate. I would like to have a number just out of curiosity.

So I've got an ellipse. This may take a while though because my geometry is bad. I've got an ellipse with a perimeter of about 34, and major axis is something along the lines of 18" and the minor axis is maybe...I don't know...8"... And...Oh, geez...

Yea, it's going to be very messy. So I will dispense with that, and instead make another rough estimate, and rather assume myself to be...well, I'm not going to bother to do this, since it's not very exact, anyway. But I could draw a circle, a little bit smaller than that maybe. Well that circle has got...how much... something between 8 and 18, and looking at this I guess you have to stretch and elongate it in the width more than the height...closer to 18...and say 14 in diameter. So that would mean 7" in radius. So, I've got $\pi \times 49$. And that would be my guess for that and the height would be...I don't know...about 15.

Now, I've covered the torso, the two legs, and the two arms.

Ok, for the hands. I'm going to have to make another rough estimate. If I put my hand into a fist I get a little cylinder of maybe an inch and a half and a height of about 4. So I've got two hands with a height of 4, π and the radius of 3/2.

And I have no idea what I'm going to do with my feet. Well, you could make these into little rectangular prisms. 4 x 2 x 10. No actually that looks about right.

Well, maybe the neck, if we're going to be precise about it is going to be 4" in diameter, so we've got a 2" radius neck. So that would be 4π in area, in volume of it. Yea, $4\pi r^2$. And now I would add all these up. Do I have to add them up too?

S: We'll just call that number capital N, and then I'll get Mr. Knop's calculator and we'll actually do it out of curiosity.

Ok, the number is N. Ok, now that I've got the volume of a body, now I've got to figure out what the volume of a cell might be.

And it seems to me something along the lines (unclear). The diameter of a hydrogen atom is like an angstrom unit, and that's something like ten to the minus ten cm. And that's not going to be anything close to the size of a cell. So, if I had to go with the size of a cell...this is a very rough estimate, it might not even be in the right magnitude...it should be 10,000 to the inch or 10,000 cells to the cm. Maybe I'll make a compromise and say 100,000 cells to the inch is right. So that would give me 10^5. So each one is 10^5 in diameter, so we should call them spheres since that would make it simpler. I would have $10^5/2^2$ times π. Is that right? $10^5/2$...you've got 10^5 to the inch so it would be ten to the negative fifth inches over two for the radius...so square that and multiply by π. So you take that and divide it by π.

And I'm going to say that that should give you the volume, but somehow I'm not convinced that that's the case. Well, maybe it would be right because you're going to have a ten to the minus ten in the denominator there, and you multiply these things are going to come out to a good 1000 or so. So hopefully a couple thousand square inches or so when you multiply it...

The student was told that he had computed the area of a circle rather than the volume of a sphere. He made the correction, and then computed all the volumes with the help of a calculator (to 4 place accuracy before rounding off).

Appendix 2*

The Full Analysis of a Protocol

Appendix 3 presents the full protocol of two students working on the following problem:

> Consider the set of all triangles whose perimeter is a fixed number, P. Of these, which has the largest area? Justify your answer as best you can.

Student K is the same student that appeared in protocol 1. Student D (not the same as student D in protocol 2) was a freshmen with one semester of calculus behind him. This protocol was taken at the end of my problem-solving course, while protocols 1 and 2 were taken at the beginning.

The parsing of protocol 3 is given in Figure 1. The analysis given below follows that parsing.

Insert Figure 1 about here

Episode 1 (Reading, items 1, 2)

a. The conditions were noted, explicitly.

b. The goal state was noted, but somewhat carelessly (items 10, 11).

c. There were no assessments, simply a jump into exploration.

Transition 1 (Null)

a, b, c, d. There were no serious assessments of either current knowledge or of directions to come. These might have been costly, but were not--assessments did come in E_2.

Episode 2 (Exploration, items 3-17)

a. The explorations seemed vaguely goal-driven.

b. The actions seemed unfocused.

*Taken from Schoenfeld (in press).

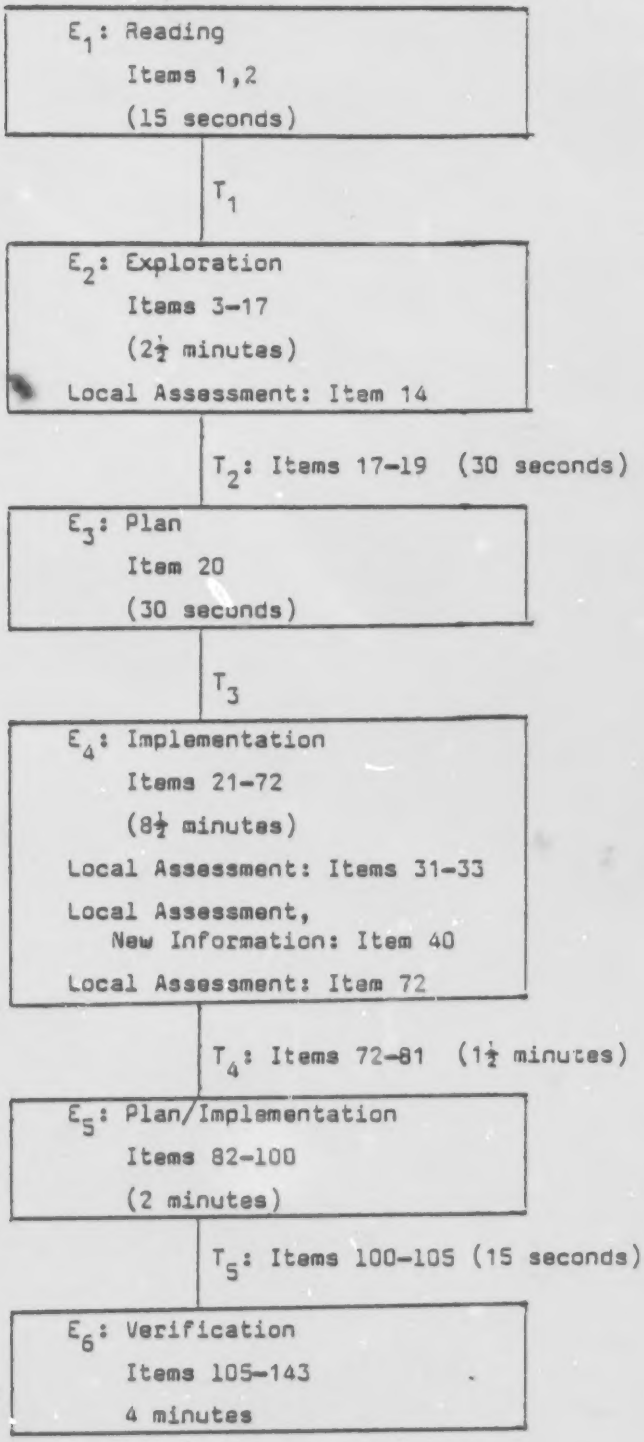


Figure 1

A Parsing of Protocol 3

c, d. There was monitoring, at items 14-17. This grounded the explorations, and led into Transition 2.

Transition 2 (Items 17-19)

a, b, c, d. Assessments were made both of what the students knew, and of the utility of the conjecture they made. The result was the establishment of a major direction: try to prove that the equilateral triangle has the desired property, and of a plan (episode 3). NOTE: If this seems inconsequential, contrast this behavior with the transition T_1 in protocol 1. The lack of assessment there, in virtually identical circumstances, sent the students on a 20 minute wild goose chase!

Episode 3 (Plan, item 20)

a. The plan is overt.

b. It is relevant and well structured. As to appropriateness and assessment, see the discussion of T_3.

Transition 3 (Null)

a, b. There was little of value preceding the plan in item 20; the questions are moot.

c. There was no assessment of the plan; there was immediate implementation.

d. The plan was relevant but only dealt with half of the problem: showing the largest isosceles was the equilateral. The "other half" is to show that the largest triangle must be isosceles, without which this part of the solution is worthless. . .a point realized somewhat in item 72, 8 minutes later. The result was a good deal of wasted effort. The entire solution was not sabotaged, however, because monitoring and feedback mechanisms caused the termination of the implementation episode (see the sequel).

Episode 4 (Implementation, items 21-72)

a. Implementation followed the lines set out in episode 3, albeit in somewhat careless form. The conditions were somewhat muddled as the first differentiation was set up. The next two local assessments corrected for that (better late than never).

Local Assessment (Items 31-33)

1, 2, 3. The physically unrealistic answer caused a closer look at the conditions--but not yet a global reassessment (possibly not called for yet).

Local Assessment, New Information (Item 40)

1, 2, 3. The "new information" here was the realization that one of the problem conditions had been omitted from their implementation ("we don't set any conditions--we're leaving P out of that"). This sent them back to the original plan, without global assessment. The cost: squandered energy until item 72.

Local/Global Assessment (Item 72)

This closes E_4. See T_4.

Transition 4 (Items 72-81)

a, b. The previous episode was abandoned, reasonably. The goal of that episode, "show it's the equilateral," remained. This, too, was reasonable.

c, d. They ease into Episode 5 in item 82. (It's difficult to say how reasonable this is. Had they chosen something that didn't work, it might have been considered meandering. But what they chose did work.)

Episode 5 (Plan/Implementation, items 82-100)

a, b. "Set our base equal to something" is an obviously relevant heuristic.

c. They plunge ahead as usual.

d. The variational argument evolved in a semmingly natural way.

e. There was local assessment (item 95). That led to a rehearsal of the sub-argument (item 96), from which D apparently "saw" the rest of the solution. Further (item 100), D assesses the quality of the solution and his confidence in the result.

Transition 5 (Items 100-105)

a, b, c, d. The sequel is most likely the result of a two-person dialectic. It appears that D was content with his solution (perhaps prematurely), although his clarity in explaining his argument in E_6 suggests he may have been justified.

Episode 6 (Verification, items 105-143)

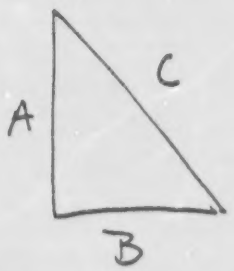
This is not a verification episode in the usual sense. K's unwillingness to rest until he understands forced D into a full rehearsal of the argument and a detailed explanation, the result being that they are both content with the (correct) solution.

Appendix 3

Protocol 3

1. K: (Reads problem.) Consider the set of all triangles whose perimeter is a fixed number, P. Of these, which has the largest area? Justify your assertion as best you can. All right now what do we do?

2. D: We got a triangle--well we know we label sides A, B and C.



3. K: Right. I'll make it a right triangle--all right--A,B, C and the relationship such as that 1/2AB = Area and A+B+C = P and $A^2 + B^2 = C^2$ and somehow you've got an area of one of these in the perimeter.

4. D: Yeah,except for somehow--I mean I don't really know--but I doubt that's the triangle of minimum area--well, o.k. we'll try it.

$\frac{1}{2}AB = A$

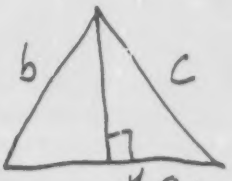
$A+B+C=P$

$A^2+B^2=C^2$

5. K: Largest area. Well, it is the only way we can figure out the area.

6. D: All right.

7. K: But for an isosceles we can do almost the same thing. This is 1/2(A). So that we know that the area is $(A/2)\sqrt{C^2-(A/2)^2}$. The perimeter = A + B + C and the height equals $\sqrt{C^2-(A/2)^2}$.



$b=C$

8. D: All right.

9. K: Now what do we do. We've got to figure out the largest area.

10. D: Isn't it the minimum?

11. K: The largest area.

$A=\frac{1}{2}(a)\sqrt{c^2-(\frac{1}{2}a)^2}$

12. D: So actually if we can get A--we have to get everything in terms of one variable and take the derivative, right? Basically?

$a+b+c=P$

13. K: Yeah,well--

$h=\sqrt{c^2-(\frac{a}{2})^2}$

14. D: Well, I still don't know if we should do--I mean we can find an area for this and can find an area for that, granted, but if we

ever come to a problem like this--I mean we don't know--we have no idea as of yet with a given perimeter what's going to be that.

15. K: Right.

16. D: So, there--I mean--you can do that again but then what do you do?

17. K: Then we're stuck, right? Usually, you know, you could probably take a guess as to what kind of triangle it would be--like you could say it is a right triangle or an isosceles--I think it is an equilateral, but I don't know how to prove it.

18. D: Umma.

19. K: So we have to figure out some way to try to prove that.

20. D: All right, a good guess is that it is an equilateral, then why don't we try an isosceles and if we can find that these two sides have to be equal to form the maximum area, then we can find that--then we should be able to prove that side also has to be equal.

21. K: O.k. so B will be equal to C, so the perimeter P = A + 2B, or A + 2C = P.

22. D: All right.

23. K: Ummmm.

24. D: See what we've got.

25. K: Fix A as a constant then we can do this, solve that for C.

26. D: All right.

27. K: For a maximum area we've got 1/2, let's say A = 1, $)C^2 - 1/4$, right? Maximum area: $1/2(C^2-1/4)^{1/2} = 0$.

28. D: C^2 minus what?

29. K: $(1/2)^2$, yeah, $(1/2)^2$. A/2, where A = 1. O.k.?

$1/2\sqrt{C^2-1/4}$

$1/2(C^2-1/4)^{1/2}=0$

$1/4(C^2-1/4)^{-1/2}(2C)=0$

$2C=0$

$C=0$

30. D: Ah, ah.

31. K: Mumbling--this is $1/4(C^2-1/4)^{-1/2}$. 2C, so we know that 2C has to = 0 and C = 0 and we are stuck!

32. D: We should have taken a derivative in it and everything, you think?

33. K: Yeah, that's the derivative of that. So does it help us? My calculus doesn't seem to work anymore.

34. D: The thing is--pause--you are letting C be the variable, holding A constant. So what was your formula--1/2 base times square root.

35. K: The base A times the square root times the height which is a right triangle to an isosceles which is --so it is $C^2-(A/2)^2$ which would give you this height.

36. D: $A^{2/4}$, no, $A^{2/2}$, no, $(A/2)^2$.

37. K: How about P = , ... no, C = P -A/2? Should we try that--

38. D: No, see part of the thing is, I think that for here we're just saying we have a triangle, an isosceles triangle, what is going to be the largest area? Largest area.

39. K: Largest area--set its derivative equal to 0.

40. D: All right. Well the largest area or the smallest area--I mean--if we are going to take a derivative--I mean--what's going to happen is you have a base and it's going to go down like that--I mean--we don't set any conditions--we're leaving P out of that.

41. K: Ah, ah.

42. D: That's absolutely what we have to stick in.

43. K: We've got C and a P-A over 2.

44. D: P -A over 2.

45. K: Formula--isosceles.

46. D: A + 2B = P--all right?

47. K: Shall we try that--mumbling. -A over 2--we've got to have a minus 1/4 PA--

48. D: Well, then you can put A back in--then you can have everything in terms of A, right? Using this formula, we have the area and we have a--

49. K: All right--P--so that's $A/2\left(\frac{P^2-2A+A^2}{4}-\frac{A^2}{4}\right)^{1/2}$ and that's $A/2\left(\frac{P^2-2A}{4}\right)^{1/2}$...(mumbling and figuring)

50. D: Wait a minute--you just took the derivative of this right here?

51. K: This times the derivative of this plus this times the derivative of this.

52. D: Oh.

53. K: Mumbling and figuring...$A/4\left(\frac{P^2-2A}{4}\right)^{-1/2}(2P-2)+\left(\frac{P^2-2A}{4}\right)^{1/2}$

$1/2 = 0$...so $\frac{2AP-2A}{4}+\frac{P^2-2A}{8}=0$.

54. D: So can we get A in terms of P?

55. K: P^2--

56. D: $8P^2 - 8P^2$ bring the P^2 on this side and multiply it by 8 and we'll have a qudratic in terms--no we won't--then we can just have A we can factor out in the equation--you see.

57. K: O.k. $P^2 =$

58. D: $-8P^2$--oh, are we going to bring everything else to the other side?

59. K: Yeah, 2A- +4A - 4AP x 8--No--

60. D: That's not right. Well, the 8 we can just multiply--

61. K: P^2 = all this.

62. D: Right.

63. K: P^2 - 4AP = --this isn't getting us anywhere.

64. D: P^2 = factor out the A--then we can get A in terms of P.

65. K: $P^2 = 2A$--so you've got $A = \frac{P^2}{6+4P}$ --

66. D: So if we have an isosceles triangle and A has = to--

67. K: be equal to that--

68. D: And if A has to be equal to that and B and C are equal--

69. K: So, B = --(whistles)

70. D: B = P- that.

71. K: 2B = P-A over 2.

72. D: No we aren't getting anything here--we're just getting--thing is that we assumed B to be equal to C so of course, I mean--that doesn't--we want to find out if B is going to be equal to C and we have a certain base--let's start all over, and forget about this. All right, another triangle. Certain altitude.

73. K: Well, let's try to assume that it is an equilateral.

74. D: All right.

75. K: Sides--mumbling--perimeter equals 3S, right?

76. D: Yeah, but wait a minute--that's still not going to really help us--what are we going to do--simply assume that it is an equilateral. We're just going to get that it is an equilateral, of course it is going to be an equilateral if we assume that.

77. K: True.

78. D: We want to prove that it is an equilateral if we think it is. If we want to do anything we can--

79. K: Yeah, how do you prove it?

80. D: Well, we can make up a perimeter--we don't need a perimeter P, do we? So,--

81. K: Where are you going to get area formula in the form of P?

$P = 2a + 4a + 2P$

$P = 2a(3+2P)$

$\frac{P^2}{3+2P} = 2a$

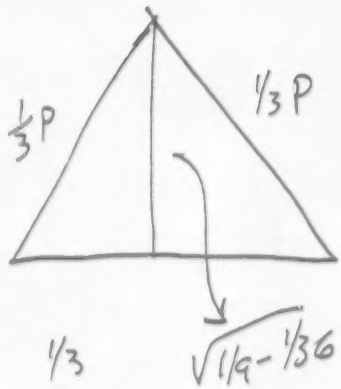
$a = \frac{P^2}{6+4P}$

82. D: We want to maximize the area so that we can prove--o.k. we have the given base--we'll set our base equal to something.

83. K: Yeah, mumbling, P, or something--I don't know.

84. D: Then the other two sides have to add up to P.

85 K: We--how about we say--let's start with an equilateral, just for the hell of it--see what happens. You get 1/3P, 1/3P and 1/3P. And this is 1/9 - 1/36 which is the height--



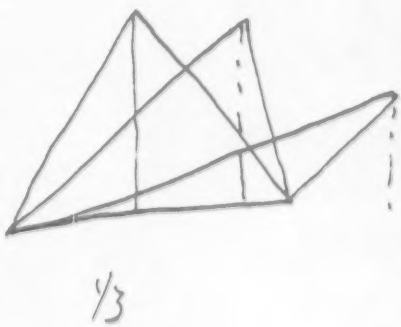
86. D: Now the thing we want to do is say--o.k. if we shorten this side at all and then what's going to happen to the height--if we leave this the same.

87. K: We can't shorten it.

88. D: And we shorten this side--sure we can--

89. K: Well--

90. D: We can have a--this equal to 1/3 and then a--this equal to--well you're going to have--I mean--



91. K: Aha.

92. D: This is going to get longer like that. Now we can see from this that all that is going to happen is that the base is going to get shorter so we know from that as far as leaving the base constant goes if we move--if we shorten this side then it is going to--somehow the point's going to go down in either direction.

93. K: Semicircle.

94. D: Right. That proves that we have to have an equilateral.

95. K: No, it proves an isosceles.

96. D: No, isosceles, I mean. All right from that if we set--we know that those two have to be equal so if we set this base equal to anything--it doesn't have to be 1/3P--we can also show that if this goes down--the area is going to get smaller, so the constant base then the height is going to get shorter and shorter and is getting smaller and smaller actually.

97. K: O.k., o.k.

98. D: In this case if it goes down to this side, we're going to have again a smaller angle here, shorter base here--and [noise].

99. K: So we get--so we know it is an equilateral--well prove it.

100. D: I don't know that's not a rigorous proof, but it is a proof--good enough for me.

101. K: Proves that an equilateral has the largest area.

102. D: Oh, we're talking about the largest area.

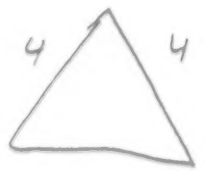
103. K: Yeah.

104. D: Oh, we just did.

105. K: We have to prove it has fixed number P--perimeter.

106. D: Well we already--we assumed that we have a fixed P, all right? I mean this is a proof as far as I.

107. K: Well, we've shown that an equilateral has the largest area. We haven't shown that if you have a certain set perimeter, let's say a right triangle, with a perimeter which is the same--we will not have a larger area.



108. D: No, but we have because we have shown with the set perimeter--o.k. we know that--

109. K: Well what if we have 3, 4, 5 with an equilateral being 4, 4, 4--

110. D: 3, 4, 5 is what? Mumbling.

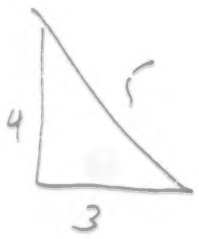
111. K: 12. So this area will be 6 and this area will be side squared 16. --o.k. that will have the largest area.

$\frac{16}{2}\sqrt{3}$

$8\sqrt{3}$

112. D: What's that 1.7?

113. K: Yeah, 8 is still greater than 6 and that's greater than 1.



114. D: Oh, yeah, that's right. Yeah, but the thing is if we have a fixed dimension, we already showed that, o.k. what is going to happen is as this

side gets longer--say we use 4 as a base here, so then what's going to happen--well say we use 3 as a base, just so we won't have an equilateral when we are done--what's going to happen as 4 gets longer and 5 gets shorter--it's going to go upwards. The optimum area--the maximum area is going to be right there. Because you've got--

115. K: Right.

116. D: This angle and that height. If you make this angle any less--maybe let me draw a picture--

117. K: I can understand that--this will give us largest area, but how can we prove this bottom is one-quarter--1/3 the area of the perimeter?

118. D: Well, remember all the problems we've done where we say--o.k. let me just start from here once more--so that we have 3, 4, 5--is that what you have--because that's going to be 5. Wasn't a very good 3, 4, 5 anyway. So you start out with 3, 4, 5--all right, we pick the 3 has the base, right?

119. K: Aha.

120. D: All right, it's 5--mumbling--if we have 3 as the base--and this is a little bit off an isosceles, but if we draw an isosceles as 3 as the base--o.k. we've got a right angle--that's got to be the maximum--mumbling--(height?) because if it goes any--

121. K: Right.

122. D: Over this way, it is going to go down.

123. K: O.k.

124. D: All right, so remember the argument we've used--well if we--

125. K: Yeah, I can show that, but what you're not showing is--what you're not proving is that--

126. D: That it has to be an equilateral?

127. K: Right. But you're not showing that this side is 1/3 the perimeter.

128. D: Right. I'm showing--first of all it has to be an isosceles. Right.

129. K: Right.

130. D: It has to be an isosceles--that means that we've got these three sides and those two are equal--right?

131. K: Umma.

132. D: Right--so now I pick this side as my base--I already picked--if that side is my base then the maximum area would have to have an isosceles--so I turn around--this side is my--

133. K: That I understand as proof, but you're not showing me that this is 1/3 the perimeter--mumbling.

134. D: If we have an isosceles triangle--if we have an equilateral triangle--then each side has to be 1/3 the perimeter--that's the whole thing about an equilateral triangle.

135. K: I know--o.k.

136. D: First we know it must be an isosceles, right?

137. K: Umma.

138. D: O.k.

139. K: I understand this.

140. D: If it is an isosceles, it must be an equilateral, right?

141. K: All right.

142. D: And if it must be an equilateral--all three sides must be equal and if the perimeter is P, all three sides must be 1/3P.

143. K: O.k. I've got it.

Beyond the Purely Cognitive:

Metacognition and Social Cognition as Driving Forces

in Intellectual Performance

Alan H. Schoenfeld
Graduate School of Education and
Department of Mathematics
The University of Rochester
Rochester, NY 14627

SE038234

Revised (April 1982) version of a paper presented at the Annual 1982 American Educational Research Association meetings, New York. To appear in Cognitive Science.

The research described in this paper was supported by a National Science Foundation grant, SED 79-19049

I. Overview

This paper is one of two whose purpose is to delineate a series of psychological and methodological issues related to the use of verbal methods (clinical interviews and protocol analyses) for research into human problem solving processes. Both papers are based on the same foundation, the premise that "purely cognitive" behavior is extremely rare, and that what is often taken for pure cognition is actually shaped -- if not distorted -- by a variety of factors. The companion paper (note 1) discusses a number of variables that affect the generation and interpretation of verbal data, for example the number of persons solving a problem, the nature of the instructions to verbalize, and how comfortable the subject feels in the experimental environment. This paper tries to place such methodologies in a much broader context, in an attempt to explicate some of the "driving forces" that generate the behaviors that we see. Briefly stated, the idea considered here is that the cognitive behaviors we customarily study in experimental fashion take place within, and are shaped by, a broad social-cognitive and metacognitive matrix. That is, the tangible cognitive actions that we observe are often the result of consciously or unconsciously held beliefs about (a) the task at hand, (b) the social environment within which the task takes place, and (c) the individual problem solver's perception of self and his or her relation to the task and the environment. It is argued that the behaviors we see must be interpreted in that light.

This is an exploratory discussion, an attempt to characterize some of the dimensions of the matrix within which pure cognitions reside. The discussion takes place in two parts. The first part outlines three qualitatively different levels of analysis that I think may be necessary to fully

make sense of verbal data, even when one's intentions are "purely cognitive." These levels are described in section 2, and a brief analysis of some protocols from that perspective is then given in section 3. In the second part (sections 4 and 5) the discussion is broadened and I try to flesh out some of the dimensions of the matrix. Much of what follows is highly speculative, and a good deal of the "evidence" anecdotal. The idea is to point out some of the pitfalls in current lines of inquiry, and to map out (one hopes) some useful directions for future inquiry.

2. Background; a Framework

I wish to suggest here that three separate levels or types of analysis may be necessary in order to obtain an accurate characterization of subjects' problem solving performance from the analysis of "verbal data" that they produce while solving problems.* These are:

A. An analysis of "tactical" knowledge. This includes the facts, procedures, domain-specific knowledge, and "local" heuristics accessible to the individual.

B. An analysis of "control" knowledge and behavior, including "strategic" or "executive" behavior and conscious metacognitive knowledge.

C. An analysis of consciously and unconsciously held belief systems, and the way that they "drive" problem solving behavior.

*There are, of course, many levels of analysis beyond those discussed here. At the microscopic level, see Monsell's [1981] review of what he calls the "nuts and bolts of cognition:" representations, processes, and memory mechanisms. At the very macroscopic level, there is the broad set of social cooperative behaviors within which "real" problem solving actions often take place. "Real world" problem solving, too, is beyond the scope of this study. Here we shall focus on analyzing the protocols obtained from students under relatively ideal laboratory situations.

Each of these categories is described below. As background, however, it is important to characterize some of the defining properties of the first two categories, "tactical" and "strategic" knowledge and decisions. Roughly, the distinction is as follows. A strategic decision is a global choice, one that in a substantive way affects the direction of a problem solution and the allocation of resources (including time) to be used in a solution. Such "control" decisions include selecting goals and deciding to pursue or abandon particular (large-scale) courses of action. In short, they are decisions about what to do in a solution. In contrast, tactical knowledge and procedures are used to implement the strategic decisions. They deal with how to do what has been decided at the strategic level. Suppose, for example, that a student working on a problem decides to calculate the area of a particular region, or to "look at an easier related problem." If doing so will occupy, say, five or more of the allotted twenty minutes for solving the problem, that decision is strategic: it, alone, may "make or break" the solution. On the other hand, the decisions regarding how to implement that choice -- for example, whether to calculate the dimensions of the region by trigonometry or analytic geometry, or which easier related problem to explore -- are tactical. Note that in the latter case, the implementation of a problem solving heuristic is considered a tactical matter. This is non-standard. Some elaboration of the three categories follows.

A. On Tactical Knowledge

This category is quite broad, including as subcategories the range of facts and procedures that

are available to the individual for implementation in a problem solution. A characterization of many of the relevant issues is given in Simon's (1979) review article, "Information processing models of cognition." Simon is primarily concerned with psychological and AI simulations of expert problem solving performance in semantically rich domains. He describes the key issues as follows. "The central research questions are two: (a) how much knowledge does an expert or professional in the domain have stored in LTM [long term memory], and (b) how is that knowledge organized and accessed so that it can be brought to bear on specific problems?" The focus here is somewhat different since we are interested in analyzing students' performance to determine sources of both success and failure. But many of the issues are the same.

To begin with, one needs to know what domain-specific knowledge is accessible to the problem solver. If a student is solving a straightedge-and-compass construction problem from plane geometry, for example, (see protocols 1 and 2) does he or she know that the radius of a circle is perpendicular to the tangent line at the point of tangency? Whether the student chooses to use that fact is another matter, to be discussed later. But (obviously) a solution that depends on that particular piece of knowledge may evolve in radically different ways if the student does or does not have it, and an evaluation of the solution depends on an adequate characterization of the knowledge base. Similar comments apply to procedures relevant for the solution of a problem. In the example just cited, does the student know how to construct a perpendicular to a given line through a given point? If the student does not recall the construction, does he or she know that it can be done, so that deriving the construction is a

possibility? Or must that too be discovered? These factors determine the potential evolution, and characterization, of a problem solving session.

After the question of the posession of factual and procedural knowledge comes the question of access to it. The student may know that similar triangles have certain properites, for example, but will the student "see" or even look for similar triangles in a particular circumstance? Much "expert" performance in given domains is attributed to the posession of certain problem solving schemata; this is, indeed, the foundation of much AI research. Questions of how to represent such "compiled" knowledge are open. Among the approaches to representation "particularly worth describing [are] the predicate calculus, production systems, semantic networks,and frames" (Walker, 1981). All of these approaches take as given that there are certain regularities in experts' perceptions of problem situations, and of appropriate behavior in them. This perspective is substantiated in various ways in the literature, for example with experimental results that experts in physics (Chi, Feltovich, and Glaser, 1981) and mathematics (Schoenfeld and Herrmann, in press) see through the "surface structure" of problems to perceive "deep structure" similarities and approach the problems accordingly. Moreover, students develop problem schemata that may or may not be consistent with those of experts (Hinsley, Hayes, and Simon, 1977; Silver, 1979), and these schemata change with experience (Schoenfeld and Herrmann, in press). For a characterization of the role of schemata in students' mathematical problem solving performance, see Silver (in press).

There is yet one more level of tactical behavior, that of implementing certain problem solving heuristics. Examples of these will be seen in protocols 1 and 2. In a sense, these are nearly on a par with domain-

specific schemata. For example, "it is useful to assume that one has the desired object and then to determine the properties it must have" is a heuristic typically valuable in straightedge-and-compass constructions. Its domain-specific implementation (draw the figure and see what properties it has) is quite similar to the implementation of domain-specific schemata, such as "look for congruent triangles when faced with a problem of this nature." These heuristics, like the other categories of knowledge described above, fall into the category of tools potentially accessible to the problem solver. An inventory of these tools provides a characterization of what the problem solver might be able to use in approaching a problem. Which of these tools are selected or discarded, how such decisions are made, and what the impact of such decisions on the problem solving process is, is the next level of analysis.

B. On "Control" Knowledge and Behavior

Two students, trying to determine the characteristics of the largest triangle that can be inscribed in a given circle, guess that the equilateral is the desired triangle and set out to calculate its area. They get enmeshed in calculations and, when the 20-minute videocassette recording their performance runs out of tape, are still calculating. Asked what good the answer will do them, they cannot say. This is an extreme (although not atypical) example of what might be called an "executive" or "control" malfunction: one bad decision, unmonitored and unchecked, dooms an entire solution to failure. What the students actually knew, and what they might have done given the opportunity to employ that knowledge, becomes a moot question. In contrast an expert working on an unfamiliar problem generates a dozen potential "wild goose chases," but rejects all of them after

brief consideration. With some clumsiness, he solves a problem the students did not -- although he began working on the problem with much less domain-specific knowledge than the students "objectively" had at their disposal. It can be argued that the expert's success and the students' failure were due respectively to the presence and absence of productive "metacognitive" behaviors (Schoenfeld, in press).

One of the early researchers to stress tne importance of metacognition as a major factor in cognitive performance, Flavell (1976, p. 232) characterized it as follows:

> I am engaging in metacognition...if I notice that I am having more trouble learning A than B; if it strikes me that I should double-check C before accepting it as a fact...metacognition refers, among other things, to the active monitoring and consequent regulation and organization of these processes to the cognitive objects on which they bear.

For the most part, research in artificial intelligence has not dealt directly with issues of metacognition as they are characterized here. This is a subtle point, since many of the terms used in metacognition overlap with those used in AI (see Brown's definition, below). But the usages differ. Consider, for example, skilled problem solving in physics as modeled by production systems (Larkin, McDermott, Simon, and Simon, 1980). The idea is to model competent behavior in sufficient detail to be able to select the"appropriate" behavior, a certainly enormous task. But issues of the type that humans encounter when working on such problems -- "I've been doing this for five minutes and it doesn't seem to be getting me anywhere; should I perhaps take an entirely different perspective?" -- are not the focus of such programs. They model behavior where such problematic performance is not a "problem."

Likewise, there are difficult issues of strategy selection in any

reasonably sophisticated program. But the use, for example, of "conflict resolution strategies" to determine precisely which production will "fire" when the conditions for more than one production have been met, still operates at a very different level than the one under consideration here. Few programs deal with planning and monitoring at that level, although there are many "planning" programs. One that does, and is worth singling out for special notice, is the Hayes-Roths'(1979) "opportunistic" model. Typical planning procedures call for leaving sequences of actions unspecified until one is constrained to specify their order, and checking for conflicts when one does so. A standard example is Sacerdoti's (1977) task, "paint the ladder and the ceiling." If one tries to proceed in that order, painting the ladder precludes painting the ceiling. "Planning" means specifying actions in efficient temporal order. Sacerdoti's "nets of actions hierarchies" are designed to allow for fleshing out plans in such a way that such impasses are avoided. This whole perspective, however, assumes that one works in domains where plans are there to be "fleshed out" -- certainly not a universal condition in problem solving. In contrast, the Hayes-Roths' model is many-leveled and, if it is appropriate, shifts rapidly from considerations at one level (do B before A, instead of the other way around) to another (revising the entire plan structure because of an unforseen major difficulty). This "opportunistic" model is highly structured, but also highly data-driven. It is open to the idea that one piece of new information may cause one to see everything that came before in a new light, and call for major revisions; that each piece of information, and the current state(s) of affairs must be continually evaluated and acted upon. To my knowledge, few other programs deal directly with this kind of issue.

There are, however, some programs that specifically separate what have been called "knowledge" and "tactics" here. For example, Bundy and Welham (1981) describe a technique called meta-level inference, in which

> ...inference is conducted at two levels simultaneously...The object level encodes knowledge about the facts of the domain...while the meta-level encodes control or strategic knowledge...What are the advantages of this technique?
> -The separation of factual and control information enhances the clarity of the program and makes it more modular.
> -All the power and flexibility of inference is available for controlling search (p. 189).

This perspective is at least sympathetic to the separation of "tactics" and "strategies" described above. It is an alternative to the production model system, where the decision-making resides in the nature of the productions.

There has recently been much discussion of metacognitive issues in the psychological literature. The bulk of such research has focused on metamemory (one's awareness of how he or she stores and retrieves information), and much of that work has been developmental. See Brown (1978) for an overview of the relevant literature. Research suggests that the use of self-regulation is a large component of older children's successful memory performance (Brown and DeLoache, 1978). Speaking in general, Brown (1978) describes metacognitive behaviors as "those attributed to the executive in many theories of human and machine intelligence: predicting, checking, monitoring, reality testing, and coordination and control of deliberate attempts to solve problems. I believe that these are the basic characteristics of thinking efficiently in a wide range of learning situations." One can hardly disagree. Moreover, the converse must be stressed. Just as the presence of such behaviors may promote efficient problem solving, the absence of them may doom problem solvers to failure. Discussions of

metacognition in broad problem solving domains are rare. See Schoenfeld (in press) for one attempt in that direction, and Silver (note 2) for an overview of the issue. Note that metacognitive acts are generally taken to be conscious. The sequel will argue that unconscious determinants of cognitive performance must be taken into account as well.

C. On Belief Systems

Ulric Neisser begins the article "General, Academic, and Artificial Intelligence" (1976) with the following dialogue. It was taken from Cole, Gay, Glick, and Sharp's (1971) study of cognition in a Liberian people called the Kpelle.

> Experimenter: Flumo and Yakpalo always drink cane juice (rum) together. Flumo is drinking cane juice. Is Yakpalo drinking cane juice?
>
> Subject: Flumo and Yakpalo drink cane juice together, but the time Flumo was drinking the first one Yakpalo was not there on that day.
>
> Experimenter: But I told you that Flumo and Yakpalo always drink cane juice together. One day Flumo was drinking cane juice. Was Yakpalo drinking cane juice that day?
>
> Subject: The day Flumo was drinking the cane juice Yakpalo was not there on that day.
>
> Experimenter: What is the reason?
>
> Subject: The reason is that Yakpalo went to his farm on that day and Flumo remained in town that day (Cole et. al., 1971, pp 187-188).

The point Neisser wishes to stress is that the subject's answers are intelligent, although they are not directly responsive. "The respondents do not accept a ground rule that is virtually automatic with us: 'base your answer on the terms defined by the questioner.' People who go to school (in Kpelleland or elsewhere) learn to work within the fixed limitations of this ground rule, because of the particular nature of school experience" (p. 136). There are, Neisser argues, many dimensions to "intelligence"

beyond the types of (academic) intelligence measured by IQ tests, the (artificial) intelligence modeled in computer programs, and the "purely cognitive" intelligence studied in psychological laboratories. Of course anthropologists take that as given (see, e.g. Cole, et. al., 1971, or Lave, 1980) and some cognitive scientists have urged that the range of cognitive investigations be substantially broadened (e.g. Norman, 1979).

The dialogue quoted above serves to make another point as well, one that bears directly on current methodological issues. In the dialogue we see a clash of belief systems, where the participants see the "ground rules" for their exchange in rather different ways. Were the experimenter to declare the subject "unintelligent" because he did not answer the questions as they were posed, we would argue that he missed the point: the responses must be interpreted in the context of the social environment that generated them, and not simply evaluated as "pure cognitions." I shall argue here that the same point holds in many of our methodologically "clean" laboratory studies, and that much of what we take to be "pure cognition" is often shaped by a variety of subtle but powerful factors. These factors may include the subject's response to the pressure of being recorded (resulting in a need to produce <u>something</u> for the microphone), his or her beliefs about the nature of the experimental setting (certain methods are considered "legitimate" for solving problems in a formal setting, others not), and the subject's beliefs about the nature of the discipline itself (is mathematical proof useful, for example, or a waste of time?). This network of beliefs provides the context within which verbal data are produced, and an understanding of that context is essential for the accurate interpretation of those data.

It should be clear that these comments are not meant as a blanket

a postiori challenge to the accuracy of studies that have relied upon the interpretation of verbal data. It may well be that the issue of belief systems is moot in a number of contexts -- for example, in the analysis of experts' verbal protocols for purposes of constructing artificial intelligence programs. Experimenters tend to find their subjects among their colleagues, who are generally familiar with and sympathetic to the methodologies being used for protocol collection. It is unlikely, therefore, that an unsuspected difference in belief systems between experimenter and subject will result in the misinterpretation of the verbal data. The situation may be quite different, however, when students are the source of that data and the task at hand is to interpret (in the large) what they have produced. A miscellany of examples that document this point are offered in section 4. Some less "impressive" but more typical protocols are discussed, from the perspectives at all three levels, in the next section.

3. A discussion of three problem solving protocols

Appendix 1 gives a protocol obtained from two students working on a straightedge-and-compass construction problem in plane geometry, recorded the second day of a problem solving course. The students were friends, and felt comfortable working with each other. They were both college freshmen, and had both just completed a course in first-semester calculus. They had taken the "standard" geometry courses in high school. Appendix 2 gives a protocol recorded by the same pair of students a month later, after the intensive problem-solving course. (See Schoenfeld [1982] for a brief description.) Geometric constructions were one of the topics discussed in the course. The students had read chapter 1 of Pólya's Mathematical Discovery

(1962), and worked perhaps a dozen construction problems. Appendix 3 gives a protocol obtained from a professional mathematician who had not "done" any plane geometry for a number of years. The protocols are themselves quite eloquent. The discussion is brief, serving to illustrate some of the points made in section 2. Each of the comments made here needs to be elaborated in far greater detail. A rigorous quantitative model is in the works.

I would like to begin with a general discussion of students' behavior on problems like the one given in appendix 1. From my perspective, the most telling information regarding their behavior is derived at the level of belief systems. Students' actions are shaped by their beliefs about the way that one solves geometric construction problems and about the role of "proof" in mathematical problem solving. In my experience, the following collection of beliefs about geometric construction problems is nearly universal among college freshmen who have completed at least one semester of calculus and who had, in high school, studied the "standard" 10th grade year of geometry.* This characterization reflects an almost pre-socratic, purely empiricist perspective.

a. Insight comes from very accurate drawings. The more accurate the drawing, the more likely one is to derive useful information from it.

b. Hypothetical solutions come from dominant perceptual features of the drawings. Plausible hypotheses are ranked by their simplicity

*Note that the students may not be consciously aware of holding those beliefs, in the same way that the Kpellan native quoted above may not be conscious of the "rules" that frame his discourse. In some protocols there is clear evidence (e.g. "How can I prove that? I know, I'll construct the circle."). Much of the evidence is indirect, however. A more precise statement about the belief systems is that the students' behavior is strongly consistent with the predictions of a model based on those beliefs. The model is briefly outlined here.

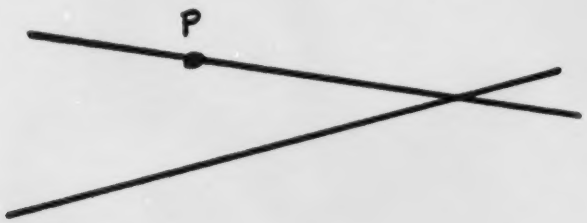
or "intuitive apprehensibility:" if you can "see your way" more clearly to the end of one plausible construction than another, the first will be ranked higher and tested first.

c. Plausible hypotheses are tested seriatum: hypothesis 1 is tested until it is accepted or rejected, then hypothesis 2, and so on.

d. Verification is purely empirical. Hypotheses about constructions are tested by performing the indicated constructions. If the construction appears to provide the desired result, then it is correct.

e. Mathematical proof is irrelevant to both the discovery and (personal) verification process. If absolutely necessary (i.e. the teacher demands it) one can probably prove that a particular construction works. But this is simply "playing by the rules of the game," verifying formally what one already knows (empirically) to be correct.

If one accepts (a) through (e) as the "ground rules" for constructions, one can predict stereotypical performance. Consider the problem given in protocol 1: Construct the circle that is tangent to the two lines in figure 1, and that has the point P as its point of tangency to one of them.



--figure 1--

Among the features of this problem that may catch the student's attention are:

F1: The radius of the desired circle is perpendicular to the top line at the point P (a recalled fact).

F2: The radius of the desired circle is perpendicular to the bottom line at the point of tangency.

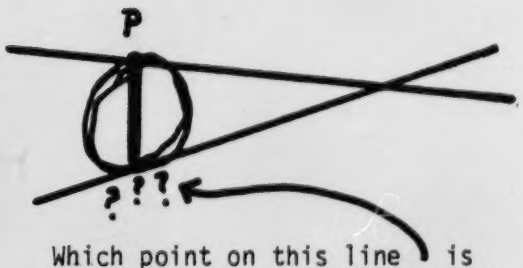
F3: By some sort of perceived symmetry, the point of tangency P' on the bottom line is directly opposite P.

F4: Any "reasonable looking" line segment originating at P and terminating on the bottom line is likely to be the diameter of the desired circle.

F5: Again by perceived symmetry, the center of the desired circle seems to be half-way between the two lines, and thus on the angle bisector.

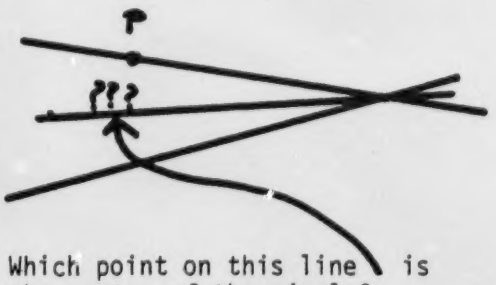
F6: The center of the circle lies on the arc swung from the vertex that passes through P.

Of these six features, F4 and F5 are perceptually dominant (and F6 is generally invoked only after F5, when one tries to identify which point on the angle bisector is the center). See figures 2a and 2b.



F4 Dominates

--figure 2a--



F5 Dominates

--figure 2b--

Various combinations of the features listed above yield hypothetical solutions to the problem. For example, F4 combines with F1, F2, and F3 respectively to generate the following hypotheses:

The diameter of the desired circle is...

H1: the line segment between the two lines that is perpendicular to P.

H2: the segment from P perpendicular to the bottom line.

H3: the segment from P to P'.

Likewise, F5 combines with F1, F2, F3, and F6 to yield the following: The center of the desired circle is at the intersection of the vertex angle bisector and...

H4: the perpendicular to P.

H5: the perpendicular from P.

H6: the segment from P to P'.

H7: the arc from the vertex that passes through P.

Finally, the non-dominant features F1, F2, and F3 combine to yield

H8: The center of the circle lies on the intersection of the perpendiculars to P and P'.

I shall argue here that the students' belief systems, as represented in (a) through (e) above, determine which of the hypotheses students will consider and in what order, and how they will test them. The following observations are major determinants of the predictions.

1. The set of candidate hypotheses is generated, of course, by the set of features that have been observed. Thus H4 through H7 can become candidates only when F5 has been noted, etc.

2. Empirically, F4 is a default condition if F5 is not noted. That is, if students do not "see" the angle bisector, they will automatically be channeled to one of H1, H2, H3, and H8.

3. H8 is less intuitively apprehensible than any of the other hypotheses, being the combination of three non-dominant features.

There are a number of other observations necessary for a complete model (for example, H5 is the least physically plausible of H4-H7), but the ones listed here suffice to indicate how it works. For example, suppose that a

stereotypical ones. This protocol is better than average (!) in a number of ways. It is relatively free of the types of pathologies described in section 4. The students work well together, and concentrate on the problem for the full twenty minutes allotted for it. Most importantly, these students demonstrate much better awareness and control of their own problem solving processes than most (see in contrast protocols 1 and 2 in Schoenfeld, in press). Their strategic and metacognitive behaviors work reasonably well -- but working within the context generated by the belief systems, these behaviors can only work to limited effect. The following is a brief running commentary.

T begins by sketching in the desired circle (Item 1), and there is a clear attempt to make sure that she and L understand the problem statement. This deliberateness in guaranteeing that they "understand" is respectable "control" behavior, in contrast to the impulsive actions taken by many students in similar circumstances.

By item 4, the sketched-in circle is erased: it was "legitimate" as an aid to understanding, but (according to their belief systems) does not belong in the figure as a proper part of working the problem. In item 5 feature F4 and the associated conjecture are introduced.

Here the dialogue is unusual in two ways. First, the students do not attend to F2 or F3, and are thus deprived of the opportunity to verify their conjecture empirically. Second, T actually raises plausible objections to the conjecture (items 6 and 8), and a meta-level dialogue ensues. This is certainly respectable executive behavior. But then the students spend 24 minutes with straightedge and compass trying to resolve the dilemma.

Their construction "looks right" (item 11) but they again recognize

that this one example does not guarantee validity in general. There is an attempt to exploit a related problem in items 14-24, again indicating some sophistication. Then five minutes (items 25-41) are spent in empirical work, resulting (finally) in the rejection of the initial hypothesis. The rejection, is, however, substantiated theoretically (the tangents to the endpoints of a diameter must be parallel).

In item 43 comes the belated recognition of F1, which again is combined with F4 to generate H1. The enthusiastic jump into implementation (items 45-50) may be in part a result of desperation, as well as the declaration that using a ruler to draw a right angle is "legal" (items 62-63). Yet items 56-57 and 61-63 say a great deal about students' perceptions of the nature of "being mathematical." Contrast this with protocol 3.

Conjecture H1 is again evaluated empirically, and the control functions are again relegated to performing post mortems: e.g. items 80-83. There is again a reference to the related problem (item 89), and -- as if we need any more evidence -- an indication that their approach to that problem was also purely empirical.* The solution degenerates from there. I wish to stress here that (a) the students did, as determined later, have an adequate factual knowledge to be able to solve the problem, and (b) their meta-level behaviors, as indicated in items 1, 6-8, 12, 14, 40-41, 80-83 and 89, are generally most respectable. The major "difficulty" is the very approach they take.

In contrast let us look briefly at protocol 3, where a mathema-

*That comment is important in the following sense. It indicates that their behavior in this experimental environment is similar to their behavior when working on the problems in their own rooms. In view of some of the examples in section 4, this is non-trivial.

tician works on the problem the students alluded to in item 14. It is, essentially, the same problem. A number of factors may contribute to the mathematician's success: better control behavior, more reliable recall of relevant facts, and (not to be underestimated) more confidence. But most important is the basic approach that the mathematician takes: he derives the information he needs tnrougn the use of proof-like procedures. Note that he is looking for congruence ("there've got to be congruent triangles in here.") long before there is a conjecture to "verify." Rather than being an afterthought or a method of verification, proof is a means of discovery for him.* The non-empirical nature of his approach is made emphatically clear the last line of the protocols, where performing the construction is the operation that is relegated to the status of an afterthought. He is certain the construction will work.

In protocol 2 we see an indication of the "intermediate" status of the students after a month of problem-solving instruction. The course focused on heuristic and executive problem solving strategies. Some of these are evident in the protocol; some were present before the course. Proof was often discussed in the course, but in the usual way: "Yes it seems that way, but how do you know it will always be true?"

Objectively the students' behavior in this protocol compares favorably with their behavior in protocol 1, along all three of the dimensions outlined in section 2. Their recall of relevant facts (e.g. that the radius of a circle is perpendicular to any tangent at the point of tangency,

*It was Pólya, I believe, who defined geometry as the art of "right reasoning on wrong figures" -- clearly the mathematician's perspective, and antithetical to the students' belief systems.

item 69) is more assured, and called into play at appropriate times. Domain-specific procedural knowledge is also more accurate, and they are confident about their abilities to perform the appropriate constructions. However, these were not disabling factors in protocol 1 and only tell a small part of the story.

There is a telling difference in their performance at the heuristic level. A few years ago that difference would have tempted me to attribute their success to the heuristics that they had learned. They draw a picture of the goal state to determine what properties it has (items 14ff.), look at extreme cases (items 34-46), consider only obtaining partial fulfillment of the conditions (item 52), and so on. The first of these heuristics alone might have guaranteed success in problem 1. However, there is a good deal more.

Their strategic (meta-level) behavior is quite good, as it was in protocol 1. They monitor and assess both the state of their knowledge and the state of the solution with some regularity (e.g. item 71), and avoid the kinds of "wild goose chases" that often guarantee failure for less sophisticated students. Here, in fact, control behaviors become a positive force in the evolution of the solution. At the very beginning (item 20), empiricism is put in its place. Time constraints are taken into account: in item 63 the expedient of using the markings on a ruler is acknowledged as "illegal" but used anyway -- they could bisect the line if they had to. They know that they are supposed to prove that their constructions "work," and predict early on that they can "do it with similar triangles and things" (item 72). In this context proof is still regarded as a means of verification, to be used after one is convinced he or she knows the

answer. The convincing comes by means of good sketches and "gut feeling," however, not by perfect constructions. "Proof by construction" is clearly put to rest in item 78.

It is tempting, then, to argue that the control strategies serve as enabling factors, allowing the students to employ their tactical knowledge with some success. Certainly the absence of efficient control behaviors would have sabotaged their attempts (Schoenfeld, in press). However, the discussion in the previous paragraph indicates that the control behaviors were operating within the context of new beliefs regarding proof and empiricism. Had those belief systems not changed, the control strategies could not have operated the way that they did. One can conjecture that without this change in belief systems their strategic behavior would still resemble their behavior in protocol 1 -- even if, say, they had been given a review of basic facts and procedures, and taken a course that stressed meta-level problem solving skills. Moreover, belief systems may affect the selection of "tactical" resources. For example, one will only select the heuristic "assume that one has the desired object and determine the properties it must have" if one believes that one can derive (prove?) useful information. The student with a purely empirical perspective will not think to implement the strategy.

This brief discussion serves merely to raise a host of questions. It is not meant to minimize the importance of tactical or strategic knowledge, but to indicate that a third and often hidden level of analysis must also be taken into account when one analyzes problem solving behavior. As indicated in section 2C, there may well be contexts in which one level of behavior predominates: the tactical in AI "expert" simulations, the strategic in "wild goose chase" solutions, and belief systems in protocol 1. Even

in this "purely cognitive" kind of investigation, other than pure cognitions must be taken into account. But this is only the beginning, as the next section indicates.

4. The Matrix Within Which Pure Cognition Resides

While the previous section raises some questions about the interpretation of verbal data, it does not at all challenge their legitimacy. That is, the discussion was predicated on the assumptions that (1) protocols like those in appendices 1 through 3 provide an accurate reflection of the cognitions and behaviors of the people who produced them, and (2) in turn, models of behavior based on such protocols (for example, the model outlined at the beginning of section 3) thus reflect the subjects' behavior with some accuracy. In the case of the particular protocols discussed, I am reasonably confident that this is the case. In general, I am much less sanguine about the "legitimacy" of verbal data, even of some data obtained in methodologically "clean" settings.

Of course this issue is not new. Methodological battles were waged, for example, over the legitimacy of introspection as a means of characterizing cognitive processes. "We have also long known, both from experiments and everyday experience, how subjects' behaviors are affected by expectation, context, and measurement procedures. The notion that there can be 'neutral' methods for gathering data has been refuted decisively" (Ericsson and Simon, 1981, p. 17). That point granted, the question then becomes one of the intrusiveness of various experimental methods. For example, it is generally acknowledged that asking subjects to analyze their problem solving processes while they work on problems does have measurable effects on performance. However, the current literature indicates that sufficiently "bland" instruc-

tions may not have a measurable effect on data gathered in the laboratory: subjects who are instructed simply to "talk out loud" as they solve problems, and not to interpret or explain, will yield essentially the same performance that they would have if they were not speaking out loud (Ericsson and Simon, 1980).

There is, in that last sentence, a very subtle but powerful disclaimer that is revealed by the following. In 1978 I made a series of recordings of students solving the following problem out loud.

> Estimate, as accurately as you can, how many cells might be in an average-sized adult human body. What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your figures?

The problem is a particular favorite of mine, an excellent task to use for examining cognitive strategies and memory searches. It can, actually, be solved without any special technical information. One wants good estimates for "average human body volume" and "average cell volume," under the assumption that there are such things. Since there will be a huge amount of guesswork on cell volume, body volume can be roughly approximated: a box with dimensions 6' x 6" x 18" will be close enough (probably within a factor of two) to the actual average.* With regard to cell size, we can see the markings of a ruler down to 1/32" so perhaps 1/50" is a lower limit to what we can see clearly without "help." Cells were discovered with early microscopes, which must have been greater than 10 power (magnifying glasses probably give about 5 power) and less than 100 power. So a "canonical cell" (say a cube) must be between 1/500" and 1/5000" on a side. The rest

*A more accurate figure can be obtained by taking an estimate of average body weight (say 150 pounds) and converting it to volume. Since the human body (barely) floats, its density is close to 1. However, the point is that there is no need to be so precise: this degree of specificity is an indulgence.

is arithmetic.

My first set of subjects were junior and senior college mathematics majors. The students knew me reasonably well and were familiar with my work. Some had done protocol recording themselves, as parts of senior projects. I took all of the appropriate precautions to set them at ease for the recording sessions, and recorded them working on the problem one at a time. See Note 1 for a representative protocol.

Typically, students would quickly choose volume as the quantity to compute. After brief consideration they would decide to compute body volume first, and would then begin extraordinarily detailed computations. Generally an "average" body (most often their own) would be approximated by a series of geometric solids whose volume was rigorously calculated. For example:

> and now a leg...a cone might be more appropriate. And the base of my leg is approximately 6 or 7 inches in diameter so you would have $(3\frac{1}{2})^2 \times \pi$ and the height would be...what is my inseam size, about 32 or 34. So you've got to have a 34 and it's a cone so you've got to multiply it by one third.

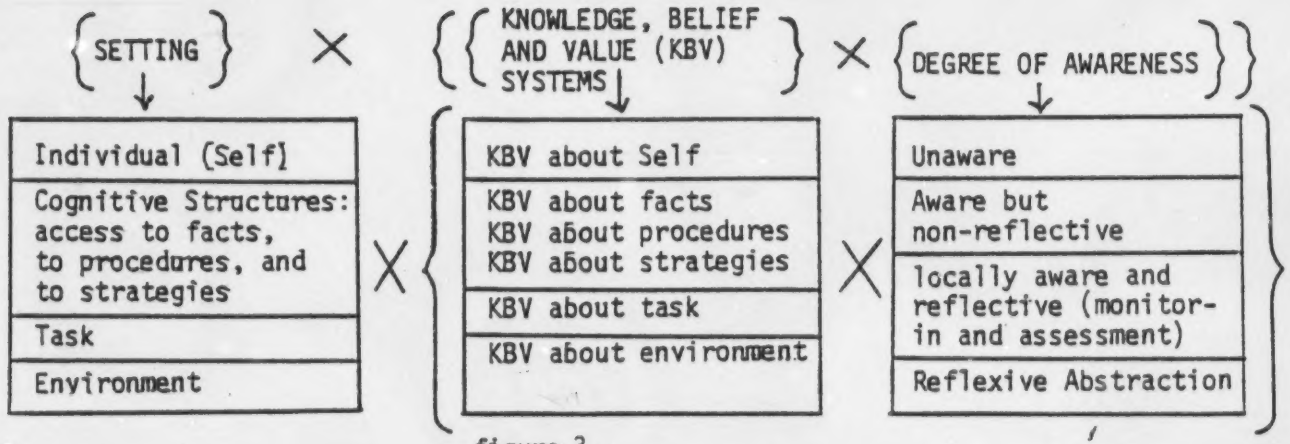
In sharp contrast to their meticulous calculations of body volumes, the students' estimates of cell size were (1) crude and (2) not accompanied by estimates of how accurate they might be. For example: "All right, I know I can see 1/16 of an inch on a ruler, so say a cell is 1/100 of an inch on a side." The students spent the great majority of their time making estimates of body volume. These results, though puzzling, were remarkably consistent.

Later in the year I began making recordings with pairs of students solving problems together. I recorded perhaps two dozen pairs of students, who solved the same problem after receiving nearly identical instructions. Not once did a pair of students demonstrate the kind of behavior I have just

described. With hindsight, it became apparent that the behavior in the single-student protocols was not a reflection of their "typical" cognitions. Rather, their behavior was pathological -- and the pathology was induced by the experimental setting itself. This problem upset the students, because they had no idea of how to approach it. Feeling on trial to produce something for a mathematics professor, they responded to the pressure by doing the only mathematics they could think of under the circumstances: computing volumes of solids. This, at least, was demonstrating mathematical behavior! (The students in two-person protocols manage to dissipate the environmental pressure between themselves, and thus to avoid extreme manifestations of pathology.)

I have dwelled on this example at length because it indicates the subtle difficulties inherent in protocol analysis. When I discovered the social causes that I now believe explain the students' behavior, I was on the verge of writing a paper describing (a) their surprising inability to make "order of magnitude" calculations, and (b) their poor allocation of strategic resources in problem solving. In hindsight this "purely cognitive" explanation of their verbal data would make no more sense than "objectively" assigning a low IQ score to the Kpellan native quoted in section 2C on the basis of his responses to the experimenter's questions. We need not travel to Liberia; clashes in belief systems between experimenter and subject occur here in our own laboratories.

Since the length of this paper has already grown out of hand, the rest of the discussion will be very brief. My intention is to sketch out some of the dimensions of the matrix within which "pure cognition" resides. A broad outline of it, given in the form of a mathematical cross product, is given in figure 3.



--figure 3--

The column on the left of figure 3 represents an "objective" description of the problem setting, the product of the two columns on the right the set of "driving forces" that operate in and on the setting. We take one column at a time.

The first column is familiar. In the best of circumstances, this is all that one need be concerned with. "Task variables" can be described objectively, and the environment as well. "Cognitive structures" are the focus of customary laboratory investigations: facts, procedures, and strategies. Under the assumption that laboratory investigations provide an accurate reflection of problem solving behavior, the investigator's focus can be on the overt manifestations of these cognitive structures. In this context the issue is more delicate: one must (somehow) ascertain the set of facts, procedures, and strategies that are potentially accessible to the problem solver.

The second column deals with belief systems. Some ideas about belief systems have reached the level of folk wisdom: for example, the notion that, through perseverance, a person will turn the belief in his or her ultimate success into self-fulfilling prophecy. A student's belief in his

or her ultimate failure will affect the verbal data one obtains as well: I have videotapes of students who never seriously engaged themselves with a problem, in order to later rationalize what they saw as their inevitable failure. (This has been admitted to me, long after taping, by more than one student.) Beliefs about the very nature of facts and procedures will determine students' performance. The student who believes that mathematical knowledge must be remembered will be stymied when a particular object (say a procedure for constructing a line parallel to a given line) is forgotten, while another who believes that the procedure can be derived will act rather differently. The effects of strategic and task-related beliefs (one approaches constructions empirically, etc.) were considered in section 3. And the effect of beliefs about the environment (one must produce mathematics when one is solving problems for a mathematics professor!) were the causes of the pathological examples that began this section. These examples barely scratch the surface, of course. But the point is that if we wish to describe behavior as it occurs, we must worry about such things.

The third column reflects the degree to which the individual is aware of his or her knowledge and belief systems. This column represents an important extension of the current literature on metacognition, which focuses on individuals' conscious "control" activities. The discussion of unconcious determinants of behavior is vital for the following reason: one can only act upon those beliefs of which one is aware. As long as the students in protocol 1 believed that discovery and proof in geometry are purely empirical, they would continue to approach problems that way. Once they were made aware of that belief (and that other possibilities exist) they could change their behavior. Similarly, students who are aware that they can monitor and assess

their own cognitive strategies can, then, serve as active agents in their own growth. Making students aware of their own (and competing) beliefs may be one of the most valuable functions we can perform as educators. In fact, it may be necessary to do so if we expect them to hear what we say in the classroom.

5. Discussion

This paper covered a huge amount of territory, much of it at break-neck speed. First let me highlight some of the methodological issues.

A. As indicated in section 2, there are at least three qualitatively different levels at which one can analyze verbal data. Depending on circumstances, one level or another may provide the "key" to understanding what happens in a given protocol. Examples of primarily "tactical" protocols are those gathered from experts working on routine tasks in familiar domains, e.g. those in Larkin, McDermott, Simon, and Simon (1980). Examples of primarily "strategic" or executive protocols are those where students go off on "wild goose chases," e.g. those in Schoenfeld (in press). An example where belief systems provide the primary level of analysis (protocol 1) was discussed in section 3. A comprehensive discussion of verbal data requires the consideration of all three levels.*

B. Belief systems can be modeled. Such models exist, for example, in decision theory. Kahneman and Tversky's (1979) prospect theory includes computational models of decision-making that take into account subjects' belief systems. The gain or loss of the same dollar amount (say $1000)

*This is oversimplified, of course. Belief systems may have served to "explain" most of protocol 1, but protocol 2 provided a (perhaps more typical) example of the dynamic interplay among the different levels. The "real" question, as I see it, is: what accounts for the differences in problem solving performance between the two tapes? This question is of nearly overwhelming complexity. This framework offers, I hope, a first step towards unraveling it.

are not viewed in the same subjective terms: generally, loss is more traumatic. Similarly, winning $2000 may not have twice the emotional value of winning $1000. Prospect theory assigns to each of the dollar amounts above its subjective value (say, for example, -1200 for the loss of $1000, +800 for the gain of $1000, and +1450 for the gain of $2000). These figures are used to make computations of "subjective expected utility," which have reasonably good predictive power.

I believe that rigorous models characterizing the effects of belief systems on problem solving behavior can be made, and that these models will have both ecological validity and predictive power. The discussion of "typical" student behavior on geometry constructions that began section 3 is, in essence, a prospectus for that kind of model.

C. Great care must be taken in the interpretation of verbal data. It may well be true that, with sufficiently bland instructions, students' performance in the laboratory may not be measurably changed by speaking "out loud" as they solve problems. But the behavior that they produce may be completely abnormal -- even if it is consistent enough to model with great accuracy. Under such circumstances, we may simply be modeling abnormal pathology in the name of cognition. Again, the issue may be moot where the belief systems of the people on both sides of the microphone coincide (with experts generating protocols for their colleagues' simulations). But the more alien the setting for the subject, the more likely it is that the data will be "driven" by covert beliefs that skew its meaning (see Note 1).

The second set of issues deals with applications of cognitive research

to educational research and development. Here the potential for the misunderstanding and misapplication of basic cognitive research is enormous. There are dangers in adapting both the methods and results of much current research to educational settings.

D. Researchers in education increasingly rely on "verbal methods" such as protocol analysis for their research, using for their analyses the successful analytical tools and perspectives derived from AI and information processing research. Yet the goals and the contexts of such studies can be substantially different. In much AI work the goal is to model idealized, purely cognitive behavior. Both the subjects and the tasks are selected to facilitate this kind of modeling, and a "purely cognitive" approach appears to be sufficient. In educational work, characterizing "idealized" intellectual behavior is only one component of a much larger enterprise. If one wishes to affect students' behavior, one must be able to describe it accurately and to characterize what causes it -- and it would appear that belief systems are a major driving force in students' behavior. Any framework that ignores them -- regardless of how accurate it is in other contexts -- can result in the severe distortion and misinterpretation of the data.

E. The applications of cognitive research to schooling must take into account the context in which cognitions are embedded. The brief discussion of figure 1 in section 4 is an attempt to sketch out the range of issues that must be taken into account if our increasing knowledge about cognition is to be employed usefully in the schools. There are any number of examples regarding that context. Jean Lave (Note 3) reports that people's use of arithmetic in everyday situations does not correlate

well with their scores on paper-and-pencil tests of it. Dick Lesh (Note 4) reports that students' problem solving behavior when dealing with "real" problems bears little or no relation to their "academic" problem solving behavior. Neisser (1976) argues the point in general.

F. The characterization of "typical" student behavior given in section 3 provides both an indictment of our current mathematical instruction and a warning about the dangers of presenting instruction that is incompatible with students' belief systems. The students from whom I gathered data were above average in a number of ways, had studied calculus, and were mathematically motivated as well: they had chosen to enroll in my problem solving course. They had taken a year of geometry in high school, and the vast majority of that time was spent in proof-related activities. Two years later, we see what remains from all that instruction (what they really learned?): a thoroughly empirical perspective diametrically opposed to mathematical proof and argumentation. When our instruction contradicts students' belief systems, it (ultimately) rolls off them like water off a duck's back. One must take students' belief systems into account in order to provide them meaningful instruction.

I think that a broad attempt to deal with cognition in its "real world" context can have a strong positive effect on schooling. The three dimensions that appear most critical to me are represented in the three columns of figure 3. It goes without saying that knowledge of the basic facts, procedures, and strategies (the first column) is essential. Most of this paper has argued for the importance of the second column, and I will not labor the point further. The third, "awareness," is worth discussing a bit more. I would assume that the purpose of schooling is to prepare

students for life after school: to help them develop the mechanisms they will use throughout life to adapt to new situations. Yet virtually all of the college freshmen in my problem solving courses enter the course completely unaware of the fact that they can observe, evaluate, and change their own behavior! It is as if their minds are autonomous, independently functioning entities, with the students as passive (oftimes frustrated) spectators. As long as this remains the case, the students are slaves to their own behavior. Once this belief, or any other, is made conscious, it can be acted upon and changed. Providing students with the potential for this kind of adaptation may be the greatest service we can render them.

Reference Notes

1. Schoenfeld, A. On the analysis of two-person problem solving protocols. Paper presented at AERA annual meetings, New York, March 1982.
2. Silver, E. Thinking about problem solving: toward an understanding of metacognitive aspects of mathematical problem solving. Paper prepared for the Conference on Thinking, Fiji, 1982.
3. Lave, J. Arithmetic procedures in everyday situations. Paper given at AERA annual meetings, New York, March 1982.
4. Lesh, R. Precesses used in applied problem solving. Paper given at AERA annual meetings, New York, March 1982.

References

Brown, A. Knowing when, where, and how to remember: a problem of metacognition. In R. Glaser (Ed.), Advances in Instructional Psychology (Vol. 1). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978.

Brown, A., and DeLoache, J.S. Skills, plans, and self-regulation. In R. Siegler (Ed.), Children's Thinking: What Develops. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978.

Bundy, A. and Welham, B. Using meta-level inference for selective application of multiple rewrite rule sets in algebraic manipulation. Artificial Intelligence. 16, (1981), 189-212.

Chi, M., Feltovich, P., and Glaser, R. Categorization and representation of physics problems by experts & novices. Cognitive Science, 5, (3), 121-152.

Cole, M., Gay, J., Glick, J., and Sharp, D. The Cultural Context of Learning and Thinking. New York: Basic Books, 1971.

Ericsson, K., and Simon, H. Verbal reports as data. Psychological Review, 1980, 87, pp 215-251

Ericsson, K. and Simon, H. Sources of evidence on cognition: a historical overview. In T. Merluzzi, C. Glass, and M. Genest (Eds.), Cognitive Assessment. New York: Guilford Press, 1981.

Hayes-Roth, B., and Hayes-Roth, F. A cognitive model of planning. Cognitive Science, 1979, 3, 275-310.

Hinsley, D., Hayes, J., and Simon, H. From words to equations: meaning and representation in algebra word problems. In P. A. Carpenter and M. A. Just (Eds.), Cognitive Processes in Comprehension. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Kahneman, D., and Tversky, A. Prospect theory. Econometrics. 47, pp 263-291

Larkin, J., McDermott, J., Simon, H., and Simon, D. Expert and novice performance in solving physics problems. Science, 1980, 108, pp. 1335-1342.

Lave, J. What's special about experiments as contexts for thinking? The Quarterly Newsletter of the Laboratory of Comparative Human Cognition. 1980, 2, pp 86-91.

Lesh, R., Mierkiewiicz, D., Kantowski, M. (Eds.) Applied Mathematical Problem Solving. Columbus, OH: ERIC, 1979.

Monsell, S. Representations, processes, memory mechanisms: the basic components of cognition. Journal of the American Society for Information Science, Sept. 1981, 32(5), pp 347-363.

Neisser, U. General, academic, and artificial intelligence. In L. Resnick (Ed.), The Nature of Intelligence. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1976.

Newell, A., and Simon, H. Human Problem Solving. New Jersey: Prentice-Hall, 1972.

Norman, D. Twelve issues for cognitive science. Report No. 7904, University of California, San Diego, Center for Human Information Processing, La Jolla, CA, 1979.

Pólya, G. Mathematical Discovery (Vol. 1). New York: Wiley, 1962.

Resnick, L., and Ford, W. The Psychology of Mathematics for Instruction. Hillsdale, New Jersey, 1981.

Sacerdoti, E. A Structure for Plans and Behavior. New York: American Elsevier, 1977.

Schoenfeld, A. Episodes and executive decisions in mathematical problem solving. In R. Lesh and M. Landau (Eds.), Acquisition of Mathematical Concepts and Processes. New York: Academic Press, in press.

Schoenfeld, A. Measures of problem solving performance and of problem solving instruction. Journal for Research in Mathematics Education, 13(1), January 1982, pp 31-49.

Schoenfeld, A. and Herrmann, D. Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology, in press.

Silver, E. Student perceptions of relatedness among mathematical verbal problems. Journal for Research in Mathematics Education, 1979, 10(3), pp 195-210.

Silver, E. Knowledge organization and mathematical problem solving. In F. Lester (Ed.), Perspectives on Research in Mathematical Problem Solving. Philadelphia: Franklin Institute Press, in press.

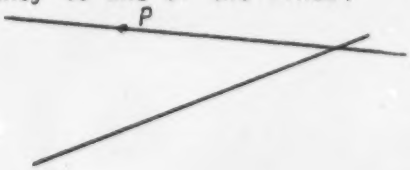
Simon, H. Information processing models of cognition. Annual Review of Psychology, 1979, 30, pp 363-396.

Walker, D. The organization and use of information: contributions of information science, computational linguistics, and artificial intelligence. Journal of the American Society for Information Science, September 1981, 32(5), pp 347-363.

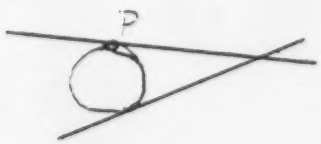
Appendix 1: Protocol 1

Problem worked the first week of instruction, by students L and T (college freshmen who had completed one semester of calculus).

You are given two intersecting straight lines, and a point P marked on one of them, as in the figure below. Show how to construct, using a straight-edge and compass, a circle which is tangent to both lines and has the point P as its point of tangency to one of the lines.



1. T: reads the problem. Oh, ok. What you want to do is that (sketches in a circle by hand), basically. Ok, how?



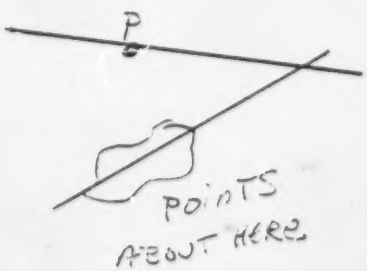
2. L: Now, ok, we have to find the center.

3. T: Of what?

(ERASES CIRCLE)

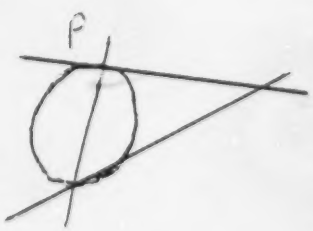
4. L: Of the circle. We are trying to find the circle, right? If we did that then we could...oh, and the radius of course.

5. T: All right, well we know the point of tangency on this line is going to be right here (points to P). What we need to find is where the point of tangency is going to be on this other line, I think. So we can find the diameter in which case we can find the center.



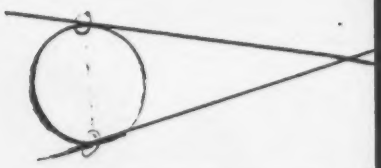
6. L: Is that...that's not necessarily true, is it? Is it true that if you have a circle like that (see right), and then that (points with finger) would be the diameter. You know what I mean? Or maybe you couldn't have it that way...

7. T: The circle has like...no, you don't have a diameter running up through there. No, we have to find the diameter from the point of tangency on this line to the point of tangency on this line, wherever it lies.



8. L: No, wait: the point of tangency, the point of tangency here, would the line connecting those two points be the diameter? It seems that you could maybe construct one where it wouldn't always work.

9. T: Wait, but see, I don't know, we're not drawing it (i.e. sketching it) the right way,



10. L: Wait, do you want to try drawing it (with the compass) and see...

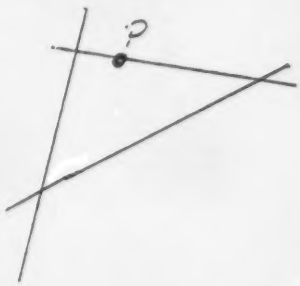
    (2½ minutes elapse in empirical work. 'A' reasonably accurate drawing results.)

11. L: So, maybe it looks like it might be opposite, see?

12. T: But would that be true for any triangle? Oh, but see...

13. L: I'm confused. I don't think it would be. Let's say you had your radius over here and you went like that. I don't think that could be...ok, I think there could be, there is a possibility.

14. T: Remember on the first problem sheet we had to inscribe a circle on a triangle? Could you do that? I couldn't.



15. L: I couldn't either.

16. T: We're in pretty sad shape. But just say we draw a triangle even though we don't know how to do it. We will draw a triangle anyway.

17. L: So how's that going to help?

18. T: Because we don't have to inscribe it actually. We just have to have something to help us (visualize it). (Draws an apparently arbitrary third line.)

19. L: Although...

20. T: Does that do anything?

21. L: Not at this point, I don't think. Maybe further along if we need a radius we could...but I don't think it does anything now.

22. T: We've gotta do something. With what we have, you just can't do it, right? We don't have enough lines or whatever there.

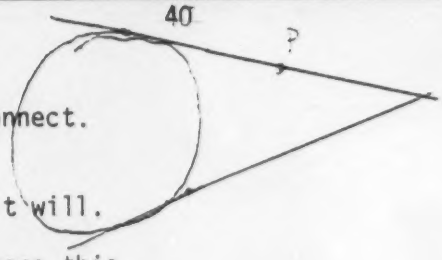
23. L: Ok, we need a center and a radius. So how do we locate the center? It has to do with, I think it has something to do with, could we do this?

24. T: No, maybe you have an equilateral triangle.

25. L: Wait, let me just try this. (Begins to expand compass.)

26. T: What are you doing?

27. L: Don't you want to see if it's true? If you have a



center way out there, because it may not connect. Don't you see? (sketch at right)

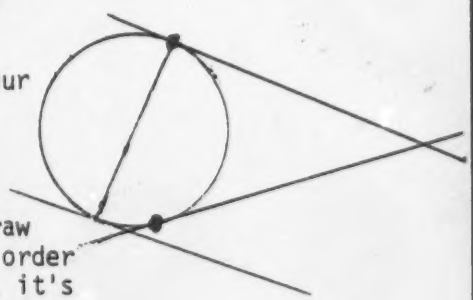
28. T: I'm pretty sure it won't. I don't think it will.

29. L: But if it won't make a circle, then that means this circle is ours (points back to earlier sketch). The one we have to deal with. You know what I mean?

30. T: I see what you mean. Like try to draw a circle out here like going through this point. See, it won't. It won't work because in order for it to work... (another few minutes with the compass. The dialogue has to do with their attempts to draw a very accurate figure, so that they can draw conclusions from it.)

31. L: Ok so that's what we're doing, right? We don't need it that big.

32. L: Yeah, wait, you couldn't because it is going to go through (the point P). I think it does have to be, right...



33. T: If we have these two points that's definitely our diameter going through it. Now we can draw...

34. L: But neither is it a tangent.

35. T: That's just what I was going to say. Can we draw these two lines so that...see you can't for in order for this to cut through this, it's too shallow, it's shallow...

36. T: Ok as soon as this...ok, make this a tangent.

37. L: In order for this to be...do you think it's going to be tangent to...

38. T: No because, because we know this one is not going to... I want to see if like we make this a tangent. You see what I mean? But that doesn't look like a diameter either. Well, I don't think that's it. Of course it couldn't be because a diameter is going to be when it's parallel, isn't it?

39. L: That's the diameter.

40. T: Ok. That's not going to help us (laughs).

41. L: You figured that out.

42. T: Right.

43. L: Can we construct one parallel to it? (Looks at original diagram.) But then we still don't know

the center.

(pause)

Could we just draw a perpendicular?

44. T: Yeah, that's what I was just going to say. If we draw a perpendicular line to this and just call that the diameter it will work from there. And then it should touch if it's perpendicular. It should be tangent at one point, shouldn't it?

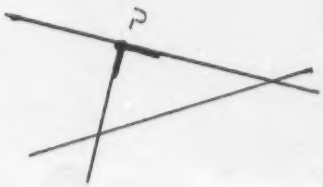
45. L: Right!

46. T: Shouldn't it?

47. L: Yes!

48. T: Won't it?

49. L: Yes!



50. T: Ok, draw a perpendicular, oh good.

51. L: Does one know how to do that with a compass? Do you?

52. T: This is a right angle, so...(uses the corner of the ruler).

53. L: Ok, that's perpendicular, ok. Doesn't look it but it is.

54. T: That's our diameter.

55. L: So if we say this is the point of tangency...

56. T: So we can bisect this to find the center, right? So call it center C. Maybe we should have done our steps.

57. L: That's all being unmathematical, completely disorganized.

58. T: Ok, back to the drawing board.

59. L: I don't know how.

60. T: Me either.

61. L: Ok, if we just use the ruler with the little numbers on it here.

62. T: Or isn't that legal?

63. L: Sure it's legal (does by hand).

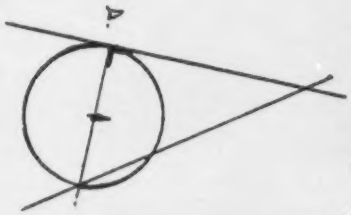
Now we have the radius, now we just draw it.

64. T: Uh, oh, do we know, we have to see if this is going to work. I know! Ugg.

65. L: My guess is, I think it's not. But we'll try.

66. T: I would think, though, it would have to, though, wouldn't it?

67. L: No.



68. T: The radius is shorter as...

69. L: I don't know. Well, let's see what happens when it goes through there.

70. T: Somehow it doesn't look perpendicular, though, doesn't it?

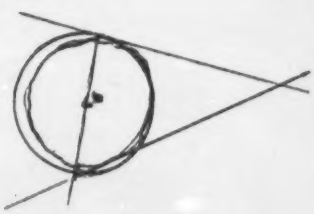
71. L: See this line isn't straight relative to the page which is why it doesn't look perpendicular.

72. T: Oh right, but...

73. L: It looks good. Now we can tell something.

74. T: Maybe, I think this tells us the point of tangency has to be way more (points to right). I think.

(Three minutes of constructions)



75. L: What circle was this one? Yup, that was a right angle. Oh, darn it.

76. T: Ok so the radius has got to be smaller because it's going outside of this line. So it's got to be a little smaller and the center has got to be up and over, like here...

77. L: But how do we...

78. T: But I don't know how to do that, without doing it until it comes out right.

79. L: Yeah.

(pause and evaluation of prior failure)

80. T: That was dumb. By doing that we were saying that no matter what this line looked like, then it looked like this, if we dropped a perpendicular we could do it and we could get the diameter for that angle and still expect to do it. You know what I mean?

81. L: Yeah, I don't think it will work for any angle though.

82. T: I know, that's what I mean.

83. L: Yeah, well, we goofed again.

(pause)

84. T: Well the only thing I can think of to do is what we did in class the other...well, what we were supposed to do, you know. The triangle thing, trying to inscribe it.

85. L: Wait, we know...

86. T: I know, that's the problem. We don't know how to do it.

87. L: I don't know what to do.

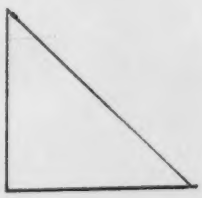
88. T: Alright, we are going to have to try something else.

89. L: Alright, what are we, what were those sort of things we tried with triangle one? Cause maybe we could... do the same thing with, on a smaller scale.

90. T: I got absolutely nowhere.

91. L: Yeah.

92. T: But I was trying to do things like, bisect this side.



93. L: Yeah, I did that.

94. T: It didn't work.

95. L: Yeah, let's see what we have here. We want to inscribe a circle in this right triangle.

96. T: Why do you want to do a right triangle?

97. L: I don't know. It just is one. Oh, I blew it now, no. The ends don't matter because we're, you see, we want to inscribe it. We're putting in the extra conditions, because it doesn't have to touch this line. It doesn't have to...oh, I don't know.

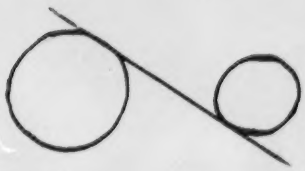
98. T: I don't think that will get us anywhere.

99. A: Ok, guys...

100. Both: We give up.

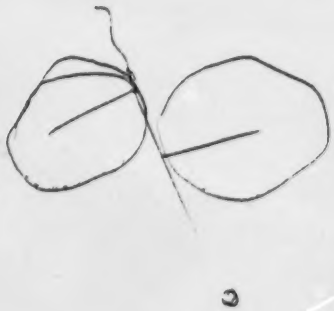
Appendix 2: Protocol 2

Problem worked after problem solving course.



The common internal tangent to two circles is the line which is tangent to both, but has one circle on each "side" of it, as in the picture to the right.

You are given three points A, B, and C as below. Using straightedge and compass, you wish to construct two circles which have the same radius, with centers A and B respectively, such that the common internal tangent to both circles passes through the point C. How do you do it? Justify.



1. T: Reads problem.
2. L: Wait, I have to read this. Ummm.
3. T: What we want basically is this, circles and a line something like this that is going to pass through here (makes sketch).
4. L: Right. Ummm.
5. T: Like that.
6. L: Except they have...where is it...have the same radius...
7. T: Uh huh
8. L: ...so it isn't going to look like that.
9. T: Right.
10. L: But, ok. Wait, I've got to think for a second.

    (erasing to draw again.)
11. L: Ok, wouldn't it...no, maybe not.
12. T: What?
13. L: No, that was dumb. Let me think.

    (pause)

14. L: Umm...should we try and draw it maybe, how it would be to see what the relationship of C is to the two circles, since that's not drawn.

15. T: Right.

16. L: You know how I am with compasses...go ahead.

17. T: Well, how big am I supposed to draw it?

(draws with a compass)

18. L: I've made this too big because they're going to overlap one another with that radius.

19. T: Yeah.

20. L: Just draw (i.e. sketch) it...you don't have to use the compass. Just draw it...just draw...no, no, no.

21. T: Ok, and I'll make my circles better. (unclear). Ok.

C

(unclear)

(good sketch)

22. T: What are you going to do?

23. L: I just want to see what it would look like more accurately (draws with compass).

24. T: Why?

25. L: Just so I could see (unclear) but you can think out loud if you have an idea. Ok. Can you think of anything? (finishes sketch)

26. T: Umm. These two radii are the same, right?

27. L: Yep. Except it doesn't look the same, does it?

28. T: That's the way you put your centers in the center.

29. L: (unclear)

30. T: (unclear) Ok. These two centers have to like... do you know what I mean?

31. L: No. Wait, what am I looking for now?

32. T: (rereads problem) Why don't we first just try to...

33. L: If we can find (unclear) (pencil placed at center point)

34. T: All right...if you just have the two centers and you go over...say the radius...the radius will have to be half way in between the centers. Alright, and then...

35. L: Say...wait...wha-wha-wha-what?

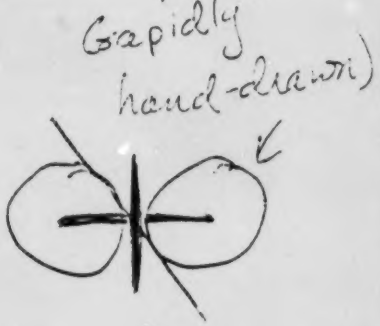
36. T: If we just try to draw the two circles and the tangent line without worrying about point C for right now.

37. L: Right.

38. T: Ok. Since they have to be of equal radius...the radius will be half way between the two centers?* It's like the tangent line would be like this.

39. L: I don't get this about the radius being half way between two centers.

40. T: Me neither.



41. L: I don't get what you mean. How's the radius half way...I don't get what you mean.

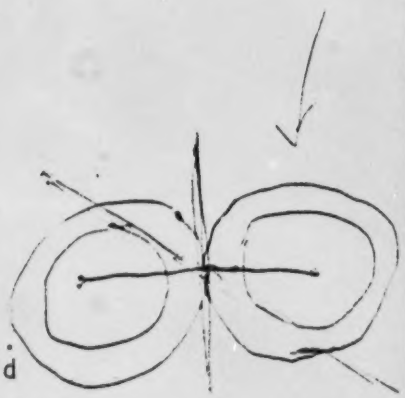
42. T: If it was like this and the tangent line would just be (unclear)

43. L: Ok, yeah.

44. T: Ok? These two have to be the same length.

45. L: Right.

46. T: And the thing that is going to determine how long they are is the angle on this line. What I mean like if they are exactly...half way in between the two centers then the line is vertical.



47. L: Right.

48. T: If we make it somehow shorter right here and here... the circles would be like this and the tangent would be on a slant like this.

49. L: Ok. Umm.

50. T: We have to figure out how they go through point C. So...

51. L: I don't know either.

52. T: Can we just start with C and draw a line through it somewhere and then make the circles tangent to it?

53. L: No.

54. T: Or...

55. L: No we're given the centers.

*She meant to say that the length of the radius in this extreme case was half the distance between the centers of the two circles.

56. T: We're also given C.

57. L: Uh huh. But just drawing the line can't guarantee you could end it with something like this if you just drew the line here. Umm. Isn't there another way we can characterize the line? Find the locus.

58. T: Ummm.

59. L: This might not work for all of them, but, look here, doesn't this look like...that's just like the center?

60. T: That's just what I was going to measure.

61. L: Umm. Because if we did that, we were given points A, B, and C.

62. T: Yes (looks at her sketch) that crosses it too. That's exactly what we're going to do.

63. L: Alright...wait, we're not allowed to use a ruler, but...yeah, divide it in half.

64. T: Yeah, bisect.

65. L: Why don't you actually do it...

66. T: Let's try it on here since we're not sure.

(Begins new sketch)

67. L: Wait, I think it was the other line. (unclear) Just connect point B. We're going to have to drop a perpendicular from B to the line.

68 T: What are you doing that for?

69. L: Because this is perpendicular and that's what the radius would be, a perpendicular and from A coming to the line also.

70. T: Right.

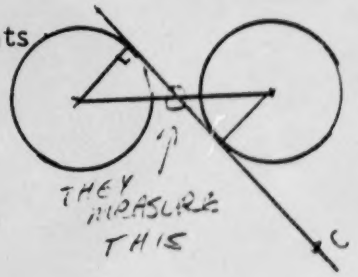
71. L: Ok. I don't know why this works, I mean, I just seem to see it, you know.

72. T: I think we can do it with similar triangles and things so let's just make sure it works (unclear).

73. L: We can do it here too...this isn't a very nice compass.

74. T: We're running out of time (whispering). Draw faster, draw faster.

75. L: I can't...this is hard.



THEIR GOOD SKETCH

(unfinished construction)

76. T: Draw faster anyway.

77. L: I didn't construct it right.

78. T: Well just draw it...it'll work.

79. L: Oh, wait, maybe I did actually. Ok, that's the radius then.

80. T: Right.

81. L: Perpendicular. Then we just have to draw...I think that's just the right thing.

82. T: That'll do it, that'll do it...wait, we've got to draw...ok, we did it. We've got to show why. We have to show that these...the reason that these are half way in between these two points is because.. angle side...we have to show that...what this side. .

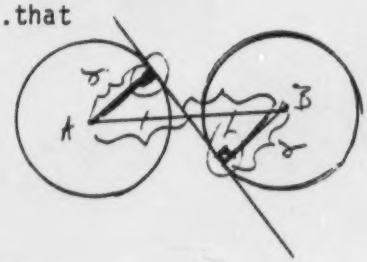
83. L: Like we have an angle.

84. T: But what are we trying to show...we want to show why this is in between A and B.

THE GOOD SKETCH

85. L: Right.

86. T: So we want to show that this is equal to this...that they...



87. Both say: ...are congruent.

88. T: Ok, we have that. We have...

89. L: ...an angle and a side. How do we know...

90. T: And we need to show that this side is compared to that side. And...

91. L: (to A): Must we prove why something works or just show you the construction?

92. A: If you can justify it I would be happy.

93. L: Ok, let's try to justify it.

94. T: Now the angle...

95. L: Well, we know, I mean, r is equal to r so it is just like...

96. T: We have these angles, so this angle equals this one.

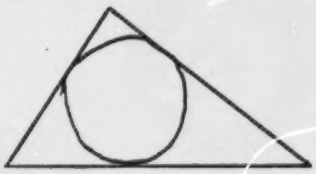
After a few minutes, and with some slight confusion, they prove that their construction has the desired properties.

Appendix 3: Protocol 3

The subject is a professional mathematician.

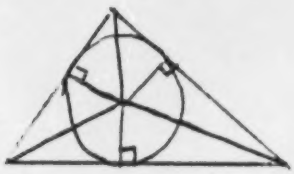
Using a straightedge and compass, inscribe a circle in the triangle below.*

(reads problem) All right, so the picture's got to look like this (draws figure) and the problem is obviously to find the center of the circle...



Now what do I know about the center? We need some lines in here. Well, the radii are perpendicular at the points of tangency, so the picture's like this (draws figure)...

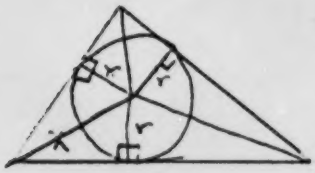
That doesn't look right, there's something missing...What if I draw in the lines from the vertices from the center? (draws figure)



That's better. There've got to be congruent triangles in here... let's see, all the radii are equal, and these are all right angles... (marks diagram) and with this, of course, this line is equal to itself (marks "x" on the figure), so these two triangles (at lower left vertex) are congruent. Great. Oops, it's angle-side-side, oh no, it's a right triangle and I can use Pythagoras or hypotenuse-leg or whatever it's called. I'm ok. So the center is on the

*The inscribed circle is a circle that lies inside the triangle and is tangent to all three sides of it.

bisectors. (Turns to investigator) I've solved it. Do you want me to do the construction?



SOME THOUGHTS ON PROBLEM SOLVING RESEARCH AND MATHEMATICS EDUCATION

Alan H. Schoenfeld
Department of Mathematics and
Graduate School of Education
The University of Rochester
Rochester, N.Y. 14627

SE 038 234

To appear in Perspectives on Research in Mathematical Problem Solving, edited by Frank Lester and published by the Franklin Institute Press.

Revised version of a paper delivered at a conference, "Issues and Directions in Mathematical Problem Solving Research," Bloomington, Indiana, May 29-30, 1981.

Part I: Issues

As I tried a number of earlier versions of this paper, I came to realize that two questions lay behind all of the issues I was grappling with:

Why do we (in particular, why do I) teach mathematics?

Why do we (I) do research in problem solving?

Since those questions look pretentious if not downright silly, I have some explaining to do. Here are some of the themes that vied for center stage in earlier versions:

Theme 1: My belief that most instruction in mathematics is, in a very real sense, deceptive and possibly fraudulent.

These are harsh words. Here are three examples to justify them.

a. "Word problems" are one of the major focal points of mathematics instruction in the elementary schools. Typical of such problems at the lower grade levels is "John had eight apples. He gave three to Mary. How many does John have left?"

Much instruction on how to solve such problems is based on the "key word" algorithm, where the student makes his choice of the appropriate arithmetic operation by looking for syntactic cues in the problem statement. For example, the word "left" in the problem given above serves to tell the student that subtraction is the appropriate operation to perform. At the research presessions to the 1980 annual NCTM meetings, the following facts were reported:

1. in a widely used elementary textbook series, 97% of the problems "solved" by the key word method would

yield the (serendipitously?) correct answer;

ii. students are drilled in the key word algorithm so well that they will use subtraction, for example, in almost any problem that contains the word "left." Problems were constructed in which the appropriate operations were addition, multiplication, and division; each used the word "left" conspicuously in its statement. A large percentage of the students subtracted. In fact, the situation was so extreme that many students chose to subtract in a problem that began "Mr. Left...."

b. I don't know about nationwide enrollment figures, but I suspect that those for Hamilton College are typical, if not low: some 60% of Hamilton's students study the calculus, but fewer than 10% of them go on to take more advanced mathematics. At the University of Rochester 85% of the freshman class takes the calculus, and more go on. Roughly, about half of our students see the calculus as their last mathematics course. Most of these students will never apply the calculus in any meaningful way (if at all!) in their studies, or in their lives. They complete their studies with the impression that they know some very sophisticated and high-powered mathematics. They can find the maxima of complicated functions, determine exponential decay, compute the volumes of surfaces of revolution, etc.

The fact is that these students know barely anything at all. The only reason they can perform with any degree of competency on their final exams is that the problems on the exams are nearly carbon copies of problems they have seen before; the students are not being asked to think, but merely applying well-rehearsed schemata for specific kinds of tasks. Tim Keiter and I studied students' abilities to deal with the pre-calculus versions of elementary word problems such as the following:

> An eight foot fence is located three feet from a building. Express the length L of the ladder which may be leaned against the building and just touch the top of the fence as a function of the distance x between the foot of the ladder and the base of the building.

We were not truly surprised to discover that only 19 of 120 attempts (four each for 30 students) yielded correct answers, or that only 65 attempts produced answers of any kind. We were surprised, however, to discover that much of the students' difficulty came not from the "problem solving" part of the process (setting up and solving systems of equations) but from the reading part of it.

> Fifty-eight protocols were obtained from randomly selected calculus students who were asked to rewrite problem statements "more understandably." Of these, 5 simply rewrote the problem verbatim. The 53 remaining rewrites tell a sorry tale: 5 (9.4%) included information which directly contradicted the input, and 11 (20.4%) contained information that was so confused as to be unintelligible. 2 students (4%) made both kinds of errors. This information is the more striking since two-thirds of these students were to write simple declarative sentences, if possible, to make their task simpler. Thus before they would normally have put pen to paper, a quarter of the 53 students had already seriously garbled or completely misinterpreted the problem statement. None of those students ever got an answer to the problem. (Keiter, note 4)

Those students had already "covered" word problems in their calculus classes.

c. I taught a problem-solving course for junior and senior mathematics majors at Berkeley in 1976. These students had already seen some remarkably sophisticated mathematics. Linear algebra and differential equations were old hat; topology, fourier transforms, and measure theory were familiar to some. I game them a straightforward theorem from plane geometry (required when I was in 10th grade). Only two of eight students made any progress on it, one of them by using arc length integrals to measure the circumference of a circle (Schoenfeld, 1979). Out of the context of normal coursework, these students could not do elementary mathematics.

In sum: all too often we focus on a narrow collection of well defined tasks and train students to execute those tasks in a routinized, if not algorithmic fashion. Then we test the students on tasks that are very close to the ones they have been taught. If they succeed on those problems, we and they congratulate each other on the fact that they have learned some powerful mathematical techniques. In fact, they may be able to use such techniques mechanically while they lack some rudimentary thinking skills. To allow them, and ourselves, to believe that they "understand" the mathematics is deceptive and fraudulent.

Theme 2: Isolationism in the Mathematics Education Community and the Relevance of Psychological Research in Problem Solving.

Mathematics education is a young and unsettled discipline. The case can be made that the phoenix of a "process oriented approach" to math-ed problem solving research rose from the ashes of the statistical approach in the mid and late 1960's; we are in our adolescence, and experiencing growing pains. Yet the community has made life much harder for itself than it had to. In a recent book on problem solving, for example, (Harvey and Romberg, 1980) five of the nine dissertation studies presented dealt with students in the fourth through seventh grades. However, the extensive literature of developmental psychology was all but ignored: a 31 page long set of references did not include a single work by Piaget. Similarly, a variety of studies in mathematics education use protocol analysis and agonize over the effects of verbalization on problem solving performance. This topic has been studied extensively in the psychological literature (Ericsson and Simon, 1978;1979). There is no need for us to reinvent that particular methodological wheel, or any of a number of others. To put it bluntly, it may be impossible to do "state of the art" work in math-ed problem solving research without a solid background in the relevant

psychological research. For example the detailed process models offered by information processing psychologists and the studies on verbalization, as well as the results on verbalization, are essential for my own work. They provide a foundation for it that is not available within mathematics education.

These comments are not meant in any way to suggest that mathematics education should become an adjunct to cognitive psychology, or even consider adopting its ideas and perspectives wholesale. It seems to me that there are significant and dangerous implications to some of the theoretical underpinnings of modern cognitive psychology, and especially of information processing psychology. I will list a few points of concern here, and discuss them at greater length below.

Among points of concern are the following. There is the phenomenon of methodologically-induced focus: one tends to examine those aspects of things that our methodologies will illuminate, and to de-emphasize or ignore those that are not compatible with them. "Models" of the problem solving process can cause difficulty in at least two ways. They may ignore aspects of the problem solving process that cannot (currently) be modeled or are incompatible with the current modeling perspective. It will be interesting to watch how information processing comes to grips with issues of metacognition, for example. Also, there is the danger that the models can be taken too seriously, as explanations of cognitive performance. In the sense just described, they may be reductive; in the sense that they are only potential explanations of performance, a particular model may be dead wrong! (See the example of the student teachers who "got the bug" in Brown and Burton (1978). We must remember that models of experts and novices are just that; the extrapolation from the models back to real people must be done with care. There is the danger that, better armed with procedures for

decomposing certain kinds of cognitive tasks, we will misuse them and become more sophisticated at perpetuating the kinds of deceptions I mentioned in theme 1. And there is currently a fair amount of (unrecognized) confusion about what it means to be an "expert" or "novice". Let me repeat a comment made to me by John Seely Brown: there can be a significant difference between "expertise" and the ability to perform well in a domain.

Theme 3: The small and possibly incestuous world of problem solving.

A few years ago I asked a number of colleagues involved in problem-solving research if they had collections of good problems to work with. Among the people who responded was Ed Silver. I was familiar with virtually all of the problems Ed sent. Most were from Polya and the other standard sources. A few were problems I had created, which (I believe) had made their way to Ed via John Lucas. The point is that the mathematics education community has a very narrow perspective on what "problem solving" means. One need only look at the 1980 NCTM Yearbook to see that virtually all the authors discuss the same kinds of "nonroutine" problems, if not the same problems themselves! (I was asked to change some of the examples in my article because they duplicated the examples in other articles.) Ed commented then that he was concerned about the incestuous nature of the community; a small number of researchers shared interests and problems, amd all seemed to be investigating this narrow collection, which went by the name of "problem solving."

I fear that his comment may be accurate. Worse, I think that we may, in teaching "heuristics," eventually become guilty of precisely the sins that offended me in theme 1: we may simply reorganize subject matter, and teach students to apply these new techniques in routine fashion. The student who has learned, in algorithmic fashion, to "substitute n=1,2,3,4 for an integer

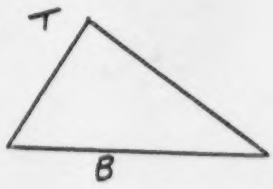
parameter and look for a pattern" may be solving difficult problems....but is he problem solving?

Theme 4: Differences between my choices of problems and my notion of "expert" from the standard choices and notions.

In a recent conversation Dick Lesh pointed out that the tasks used in my recent problem-solving studies are not the standard "non-standard" problems, and that my "experts" display markedly different (and often remarkably unproficient) behavior than most "experts." Some examples are the following.

I give both students and colleagues problems that are either unfamiliar or from domains they studied long ago. A particular favorite of mine is the following problem, although it has proved too difficult for most students:

> You are given a fixed triangle T with base B, as to the right. Show that it is always possible to construct, straightedge and compass, a straight line parallel to B such that the line divides T into two parts of equal area. Can you similarly divide T into five parts of equal area?



In fact, the problem has proven difficult for some of my colleagues. The solution provided by one GP was derided by one JTA as being stupid and clumsy. Yet I chose GP's "stupid and clumsy" solution for analysis (Schoenfeld, 1981) as an expert protocol, and found little of interest in JTA's clean solution. So my tests of problem solving do not examine what I have just taught students, and my "experts" appear unexpert by standard criteria. This is not a matter of perversity, but one of perspective. It is tied to the first three themes and to the questions with which I opened this paper. I would like to give my personal answers to the two questions, and then discuss issues 2 and 4 from that perspective.

Part 2: Questions and Personal Answers

Why do we teach mathematics? Not because mathematics is useful, although it is: our curricula reveal that. How often does one need to determine how rapidly a person could row in the absence of a current, if it takes so long to row with a constant current and so long to row against it? Or for that matter to use a trigonometric identity, virtually anything from Euclidean geometry, or to calculate the volume of a surface of revolution? Mathematics can be applied to the real world, although we do a rather poor job of teaching our students to do it. We do an even poorer job of selecting potentially useful and meaningful problems for our students to master. But that is only a part of the story.

Other parts have to do with the scope and power of the discipline. It is a massive intellectual achievement, and should be appreciated even if not used. It is as well a marvelously aesthetic discipline, and it would be nice to have our students appreciate it for that. But in my opinion the single most important reason to teach mathematics is that it is an ideal discipline for training students how to <u>think</u>. Later in this section I will try to characterize "thinking" in more detail, but for now the usual sense of the word will suffice. Mathematics is a discipline of clear and logical analysis that offers us tools to describe, abstract, and deal with the world (and later, worlds of ideas) in a coherent and intelligent fashion. Our goals as teachers should be to have students learn to use mathematics that way.

For example, the calculus version of the pre-calculus problem given above in theme 1b (Find the shortest ladder that touches the fence and the wall) is ludicrous. If one ever did need to solve such a problem, it could probably be best to do so by rough empirical methods. But it is worth having

students work on such problems. To solve this problem the student must extract the relevant information from the text, create an accurate diagram with the appropriate symbolic notation, establish goals and subgoals, and seek (from memory) the relevant information that will allow the goals and subgoals to be achieved. Further, all of this must be done with reasonable efficiency, and students must learn that as well. To the degree that this problem serves as a vehicle for developing those skills, it is worthwhile. Taken in and of itself, or as an exemplar of a class of problems, it is of questionable value. The same is true of much of the mathematics we teach.

Why do research in problem solving? From my perspective, it is so that we can better understand what constitutes productive thinking skills, so that in turn we can be more successful in teaching students to think. It is not easy to define Thinking. (I shall use the upper case T to distinguish Thinking from the ordinary associations of the word.) Here are some examples of what it is not. A mathematician is not Thinking when he uses the quadratic formula. That should come as no surprise, since the application of the formula is algorithmic. But most probably he has no need to Think when he solves the pre-calculus problem given in theme 1b. That problem is completely routine for college mathematics teachers, as are virtually all problems in the calculus. Even if he has not worked a problem isomorph of it before, the mathematician would in all likelihood be able to crank out a solution to it with as much ease as he could factor an expression like $(6x^2 + 17x + 12)$. If you were to observe (or attempt to model) his performance on that type of problem, you would be a spectator to a demonstration of domain-specific proficiency--but you would not at all be seeing whatever it is that accounts for his problem-

solving skill. The same is true for virtually all schema-driven solutions, including "heuristic" solutions to "non-routine" problems (if the "expert" has access to the schema).

To examine what accounts for expertise in problem solving, you would have to give the expert a problem for which he does not have access to a solution schema. His behavior in such circumstances is radically different from what you would see when he works on routine (including familiar "non-routine") problems. At the surface level his performance is no longer proficient; it may even seem clumsy. Lacking access to a solution schema, he has in all likelihood no clear indication of how to start. He may not fully understand the problem, and may simply "explore" it for a while until he "feels comfortable" with it. He will probably try to "match" it to familiar problems, in the hope it can be transformed into a (nearly) schema-driven solution. He will bring up a variety of plausible things: related facts, related problems, tentative approaches, etc. All of these will have to be juggled and balanced. He may make an attempt at solving it in a particular way, and then back off; he may try two or three things for a couple of minutes and then decide which to pursue. In the midst of pursuing one direction he may back off and say "that's harder than it should be" and try something else; or, after the comment, he may continue in the same direction. With luck, after some aborted attempts, he will solve the problem.

Does that make him (at least in that domain) a bad problem solver? I think not. In all likelihood someone proficient in that domain (i.e., someone who knows the right schemata) could produce a solution that puts his to shame. But that isn't the point at all. The question is: how effectively did the problem solver utilize the resources at his disposal?

One of the most impressive protocols I have ever seen is the "stupid and clumsy" solution produced by expert GP to the problem given in part 1, theme 4 (see Schoenfeld, 1981). The protocol is five single-spaced pages long (20 minutes), and a detailed analysis takes longer. GP has no idea of what "makes the problem tick," and remembers less of his plane geometry than my college freshmen, who have studied the subject much more recently. He generates enough potential sources of "wild goose chases" in his protocol to mislead an army of problem solvers. But unlike my students, he manages not to be misled. His protocol is a tour de force of metacognition: rarely do more than fifteen seconds elapse between comments on the state of his own knowledge and the state of the solution. While he is fertile in generating potential solution paths, he is also ruthless in pruning them. With less domain-specific knowledge at his disposal than most of my students had, he managed to solve a problem that left all of them stymied. Therein lies his "expertise." It is not simply the possession of schemata that allows him to solve problems with dispatch, although that is an important component of his competence--it is the ability to deploy the resources at his disposal so that he can make progress while others wander aimlessly.

Part 3: Implications

One point I wished to stress in Part 2 is that proficiency (the possession of a large number of schemata for dealing with generic classes of tasks in a domain) should not be confused with expertise. There are dangers in confusing the two.

In the short run, proficiency models (which is what virtually all "expert" models have been) are useful. It is worthwhile, for example, to develop schemata for elementary word problems that are mathematically and psychologically valid, and accessible to school children. A system of instruction based on these would obviously be preferable to the "key word" system, which uses illegitimate means to achieve what may be "rigged" performance objectives. Properly interpreted and used, Information Processing models of competent performance are valuable. In any field, "cleaner" instruction resulting in improved performance can hardly be unwelcome.

The difficulties lie, at two different levels, in how one is to interpret and use performance models. First, at the microscopic level: modeling can, at times, be an end in itself. It should instead serve as the beginning for a new set of inquiries. There are now, in a number of domains, production system models that not only simulate and predict performance but can be modified to "improve" or "grow." In some very clever work now being done at Carnegie-Mellon University (Briars and Larkin, note 1), a series of nested production models have been developed for solving elementary word problems. A running program performs at a level consistent with the performance of kindergartners. Adding one production to the system (and some minor modifications) results in performance like that of first graders, and adding one more results

in performance like that of second graders. Performance predicted by the models agrees very well with empirical data, and the models both serve to unify collections of empirical data and as a framework from which to make predictions. But do the processes in the models really reflect the cognitive processing in the children who are being modeled? The theory suffices to make predictions, without the implementation of the program. One has the feeling (in this case confirmed by conversation with the author) that the programs are important because their authors believe that the processes in the program <u>are</u>, at some level, the processes in the minds of the students. This is fine, if that attitude is considered an hypothesis to be tested. If left unquestioned, it is severely reductive and can have dangerous consequences. Moreover, it should be recognized that the hypothesis, even if correct, now gives rise to the <u>real</u> question: just what happens during a full year of a child's development that results in the addition of <u>one</u> production to his word problem "program"?

I will find it very interesting to keep an eye on this particular line of research. In a recent conversation Diane Briars told me that the students, when solving problems, would often encounter contradictions between their intuitions and the processes they had been taught to use. Students would say things like "I know it ought to be larger, but I'm supposed to use subtraction." Most often they would succumb to their training. So far as I'm concerned, the metacognitive aspects of this process--the generation of the students' intuitions, and the means that the students use to resolve the conflicts--lie at the heart of their performance. The "purely cognitive" aspects of their performance, which have been modeled, tell a critically important part of the story. However, the models do not take the metacognitions into account; they cannot, at present. To elaborate the models means to ignore an important part of

psychological reality; to deal with that reality means to abandon current methodology. Where does one go next?

These comments are not to be taken as an indictment of this study. I chose to discuss it because it is a good study, relevant to some of the themes I raised earlier. But the questions I have just raised apply to most AI studies, and are rarely raised (at least in print or in my company) by those who create them. At the recent AERA meetings Lauren Resnick characterized much AI work as "Art in the service of science." We must make certain that it does serve.

The second and much more perilous difficulty lies at the macroscopic level. There is a very serious danger when proficiency and expertise are confused, and expertise is defined to be proficiency: Thinking (with the upper case T) is then defined out of existence, or banished to irrelevancy. The situation is exacerbated by a kind of "proof is in the pudding" argument that goes something like this: "We have produced programs that operate successfully without any need for construct X. Further, people have tried to construct programs based on construct X and failed. Therefore, construct X, even if it does exist, is at best of minor importance." This particular statement was made to me about heuristics, but could also have been made about metacognition, Thinking, or any of a number of potentially important domains of inquiry. Most theoretical AI and IP work these days is done, de facto, along proficiency model lines: "experts" always seem to be performing routine tasks, and theoretical work now focuses on models of productive thinking via scripts or schemata. If the traditional evolutionary pattern holds, applied research will follow suit, and so will educational research and development. During the height of behaviorism certain "mental constructs" were déclassé, and to be shunned at all costs. Let us not make similar mistakes about Thinking in

a world dominated by proficiency models. That perspective can only deflect us from the global goals we have in teaching and research.

In sum: the mathematics education community cannot afford to ignore (as a large part of it seems to have) the psychological research on problem solving; but neither can it afford to swallow it whole. Mathematics educators have, I think, had their hearts in the right place but lacked the methodological tools that allowed for substantive and rigorous inquiries into problem solving. Many such tools have been developed by the psychological community, and much of our work will be at best second rate if we do not take advantage of them. As I mentioned above, it would be impossible for me to do my own work without the support of research into the effects of verbalization on problem-solving performance or the substantive ideas underlying the modeling of cognitive processes.

There is a great deal more to problem solving than is currently being modeled. I personally am convinced that metacognitions play a tremendous role as "driving forces" in cognitive performance, and that much more research needs to be done exploring them. They have cropped up in various ways in this paper. They include the monitoring and assessment strategies that students lack, and that allow them to go off on "wild goose chases," and that the expert has, allowing him to be efficient (Schoenfeld, 1981); they include the intuitions against which the progress or plausibility of a solution is gauged, and the means by which such conflicts are resolved; and they include both the conscious and unconscious belief systems that may determine the approaches people take to certain problems. These areas have barely been touched upon, and need much more research. They are just some of many that we will discover in an open-ended and open-minded quest for knowledge and understanding. We are beginning to make progress, and can hope to see more.

Reference Notes

1. Briars, Diane J. & Larkin, Jill H. An integrated model of skill in solving elementary word problems. Paper presented at the 1981 annual AERA meetings, Los Angeles, April 1981.

2. Ericsson, K. A. & Simon, H. A. Retrospective verbal reports as data. C.I.P. working paper No. 388. Carnegie-Mellon University,1978.

3. Ericsson, K. A. & Simon, H. A. Thinking aloud protocols as data. C.I.P. working paper No. 397. Carnegie-Mellon University, 1979.

4. Keiter, T. Solving Calculus word problems: behavior and learning. Senior Fellowship thesis, Hamilton College, 1981.

References

Brown, J. S. & Burton, R. R. Diagnostic models for procedural bugs in basic bathematical skills. Cognitive Science 2, 155-192 (1978).

Harvey, J. G. and Romberg, T. A. Problem Solving Studies in Mathematics. Madison, Wisconsin: Wisconsin Research and Development Center for individualized schooling monograph series, 1980.

National Council of Teachers of Mathematics. 1980 Yearbook: Problem Solving in School Mathematics. S. Krulik (Ed.) Reston, Va.: N.C.T.M., 1980.

Schoenfeld, A. H. Teaching problem solving in college mathematics: the elements of a theory and a report on the teaching of general problem solving skills. In R. Lesh, D. Mierkiewicz and M. Kantowski (Eds.), Applied Mathematical Problem Solving. Columbus, Ohio: Ohio State University, 1979. (ERIC Document)

Schoenfeld, A. H. Episodes and executive decisions in mathematical problem solving. In R. Lesh and M. Landau (Eds.), Acquisition of Mathematics Concepts and Processes. New York: Academic Press, in press.